



Contribution à la classification d'images satellitaires par approche variationnelle et équations aux dérivées partielles.

Christophe Samson

► To cite this version:

Christophe Samson. Contribution à la classification d'images satellitaires par approche variationnelle et équations aux dérivées partielles.. Modélisation et simulation. Université Nice Sophia Antipolis, 2000. Français. NNT: . tel-00319709

HAL Id: tel-00319709

<https://theses.hal.science/tel-00319709>

Submitted on 9 Sep 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

présentée à

l'Université de Nice-Sophia Antipolis

École Doctorale : Sciences et Technologies de l'Information et de la
Communication

pour obtenir le titre de

DOCTEUR

Spécialité : Traitement des images

Contribution à la classification d'images satellitaires par
approche variationnelle et équations aux dérivées partielles

par

Christophe SAMSON

Soutenue le 26 Septembre 2000 devant le jury composé de :

M. Olivier Faugeras	Directeur de Recherche, INRIA	Président
M. Vicent Caselles	Professor, Universitat Pompeu Fabra	Rapporteur
M. Bernard Rougé	CNES/Directeur de Recherche associé, CNRS	Rapporteur
M. Gilles Aubert	Professeur, Université de Nice	Examineur
Mme Laure Blanc-Féraud	Chargée de Recherche, CNRS	Examineur
M. Patrick Pérez	Microsoft Research/Chargé de Recherche INRIA	Examineur
Mme Josiane Zerubia	Directeur de Recherche, INRIA	Examineur

Remerciements

Je tiens à exprimer mes plus vifs remerciements à tous ceux qui ont contribué au bon déroulement de cette thèse.

Je remercie :

Les personnes qui m'ont fait l'honneur de participer au jury :

Messieurs Vicent Caselles et Bernard Rougé pour avoir accepté d'être rapporteurs de cette thèse et pour le soin qu'ils ont porté à lecture de ce document.

Monsieur Olivier Faugeras pour avoir accepté de présider ce jury.

Monsieur Patrick Pérez pour l'intérêt qu'il a porté à ce travail et pour avoir accepté de faire partie de ce jury.

Les personnes qui m'ont directement encadré tout au long de ces trois années :

Madame Josiane Zerubia pour m'avoir accueilli au sein du projet ARIANA et pour ses nombreux conseils.

Madame Laure Blanc-Féraud pour la qualité de son encadrement, pour ses conseils et pour sa disponibilité.

Monsieur Gilles Aubert pour avoir contribué, entre autre, aux aspects théoriques de ce travail et pour sa grande disponibilité.

Les "probabilistes" :

Monsieur Xavier Descombes pour les nombreuses discussions, pour ses remarques pertinentes.

Monsieur Marc Sigelle pour ses judicieuses remarques et pour son encouragement.

Monsieur Pierre Kornprobst pour les discussions que nous avons eues et ainsi que pour ses nombreux conseils avisés.

Monsieur Yoram Bresler pour m'avoir accueilli au sein de son équipe à l'Université de l'Illi-

nois à Urbana Champaign et pour ses précieux conseils.

Je tiens à remercier tout particulièrement le projet VISTA de l'IRISA Rennes, Monsieur Patrick Bouthémy son responsable, ainsi que le laboratoire COSTEL de l'Université de Rennes II, pour avoir gracieusement mis à notre disposition les différents résultats comparatifs du sixième chapitre.

Je remercie également le CNES et l'IGN pour nous avoir gracieusement fourni certaines données satellitaires et aériennes.

Je n'oublierai pas les personnes qui ont animé le projet ARIANA pendant ces trois années, à commencer par Corinne Zuzia que je remercie pour son efficacité, sa disponibilité et sa bonne humeur. Merci également à tous les "non permanents" du projet, avec une mention toute spéciale pour Radu, co-pilote co-équipier co-animateur du bureau B122, avec qui j'ai partagé pendant ces trois années, entre autre, la richesse de la langue roumaine. Norok à toi.

Pour finir, un grand merci aux amis du 33, dont Monsieur Paul Casa, pour leurs précieuses contributions.

à Patrick, parti trop tôt

Table des matières

Notations et définitions	1
Introduction	5
Chapitre 1. Position du problème	9
1.1 La problématique	11
1.1.1 La classification d'images	11
1.1.2 Présentation des données satellitaires	12
1.2 Généralités sur les approches variationnelles et par EDP	14
1.3 Quelques modèles de classification	16
1.3.1 Minimisation de distance	16
Méthodes non itératives	16
Méthode itérative	17
1.3.2 Approches stochastiques	18
Champs de Markov, inférence bayésienne, règle de Bayes	19
Modèle de Potts	21
Le chien-modèle	22
Optimisation et échantillonneurs	24
Chapitre 2. Classification avec restauration	29
2.1 Introduction	31
2.2 Restauration par modèle variationnel	31
2.3 La théorie de Van der Waals-Cahn-Hilliard	37
2.3.1 Présentation et principaux résultats	37
2.3.2 Analogies avec la classification d'images	40
2.4 Modèle proposé	42
2.4.1 Expression de la fonctionnelle	42
2.4.2 Justification	44

2.5	Minimisation et algorithme	47
2.6	Résultats expérimentaux	49
2.6.1	A propos du choix des paramètres	49
2.6.2	Données synthétiques	50
2.6.3	Données réelles	54
2.7	Remarque concernant la réponse impulsionnelle	63
2.8	Conclusion	63
Chapitre 3. Classification par modèle dynamique		65
3.1	Travaux antérieurs	67
3.1.1	Les contours actifs et formulation par ensembles de niveaux	67
3.1.2	Les modèles actifs reposant sur une approche de type région	71
3.2	La classification en tant que problème de partition	73
3.3	Formulation par ensembles de niveaux	75
3.3.1	Préliminaires	76
3.3.2	Expression de la fonctionnelle	78
3.3.3	Remarque concernant le problème de minimisation de longueur	82
3.3.4	Fonction d'arrêt, information de type contour	83
3.4	Algorithme et schéma dynamique	84
3.5	Résultats expérimentaux	85
3.5.1	A propos du choix des paramètres et de l'initialisation	85
3.5.2	Données synthétiques	87
3.5.3	Données réelles	89
3.6	Remarques	102
3.6.1	Sur le terme d'attache aux données	102
	Sur les termes globaux	102
	Remarque	103
3.6.2	Sur l'introduction d'un processus de restauration	105
3.7	Conclusion	109
Chapitre 4. Quelques comparaisons		111
4.1	Introduction	113
4.2	Modèles de comparaison	115
4.3	Comparaisons expérimentales	117
4.3.1	Présentation des résultats	117
	Image "gdr"	117
	Image "échiquier"	118

4.3.2	Commentaires	119
4.4	Conclusion	121
Chapitre 5. Le cas multispectral		129
5.1	Introduction	131
5.1.1	Le cas multispectral	131
5.1.2	Estimation du nombre de classes et de leurs paramètres	132
5.2	Extension du premier modèle	135
5.2.1	Introduction	135
5.2.2	Modèle multispectral	137
5.2.3	Minimisation et algorithme	140
5.2.4	Exemple : imagerie couleur	143
5.3	Extension du deuxième modèle	145
5.3.1	Introduction	145
5.3.2	Fonctionnelle multispectrale	145
5.3.3	Schéma algorithmique	147
5.3.4	Exemple : imagerie couleur	148
5.4	Conclusion	149
Chapitre 6. Une application sur des données multispectrales SPOT		153
6.1	Présentation des données et des modèles de comparaison	155
6.1.1	Présentation des données	155
6.1.2	Modèle markovien hiérarchique sur arbre tronqué	156
6.2	Résultats et comparaisons	158
6.2.1	Présentation des résultats	158
6.2.2	Commentaires	160
6.3	Conclusion	165
Conclusions et perspectives		169
Annexe A. Sur l'équation d'Euler-Lagrange		175
Annexe B. Premier modèle, premiers pas		177
Annexe C. Discrétisation du premier modèle		181
Annexe D. Démonstration du lemme 1		183
Annexe E. Calcul des équations (3.36)		185

Annexe F. Discrétisation du second modèle	187
Références bibliographiques	202

Liste des figures

1.1	Réflectance spectrale de quelques matériaux et positionnement des 3 bandes XS1, XS2 et XS3 du capteur SPOT.	14
1.2	Les différentes configurations associées à un modèle binaire 3×3 et leur nombres d'éléments.	22
1.3	Equations liées aux contraintes sur les contours.	23
1.4	Extension m-aire du chien-modèle.	23
1.5	Réalisations du chien-modèle (4 classes) pour différentes valeurs du triplet (e, l, n) . Simulations réalisées par Xavier Descombes.	24
2.1	Modèle de dégradation.	33
2.2	Exemple de fonctions de régularisation (voir le tableau 2.1).	35
2.3	Exemple de potentiel W à deux puits.	37
2.4	Construction du potentiel quadratique par morceaux W	44
2.5	Potentiel W pour l'image "échiquier".	51
2.6	Résultats pour l'image "échiquier". Les paramètres utilisés sont : $\lambda = 16$, $\delta = 15$, $\eta = 0.1$, avec 7 itérations sur ε	52
2.7	Classification de l'image "échiquier" à l'aide de différentes fonctions φ (voir le tableau 2.1).	53
2.8	Potentiel W pour l'image "gdr" contenant 3 classes.	56
2.9	Potentiel W pour l'image "gdr" dans le cas de 8 classes.	56
2.10	Résultat de classification pour l'image "gdr" et dans le cas de 3 classes. Les paramètres utilisés sont : $\lambda = 5$, $\delta = 10$, $\eta = 0.05$, avec 16 itérations sur ε	57
2.11	Résultat de classification pour l'image "gdr" dans le cas de 8 classes. Les paramètres utilisés sont : $\lambda = 60$, $\delta = 6$, $\eta = 0.05$, avec 40 itérations sur ε	57
2.12	Image "échiquier" : évolution du nombre de pixels mal classés (à gauche) et de la valeur du RSB (à droite) en fonction des itérations sur ε	58

2.13	Image "gdr" dans le cas de 8 classes : évolution du nombre de pixels mal classés (à gauche) et de la valeur du RSB (à droite) en fonction des itérations sur ε	58
2.14	Image "gdr" contenant 8 classes : évolution de la solution f_ε (colonne de gauche) et de la carte des discontinuités associée (colonne de droite) lors de la décroissance de ε	59
2.15	Suite de la figure 2.14.	60
2.16	Image originale SPOT (canal XS3) de la région hollandaise, et classification obtenue pour 10 classes.	61
2.17	Image SPOT originale (canal XS3) et classification obtenue pour 4 classes.	61
2.18	Seconde image SPOT contenant 4 classes. Evolution de la solution f_ε (colonne de gauche) et de la carte des discontinuités associée au cours des itérations sur la décroissance de ε	62
2.19	Image synthétique floue et bruitée (RSB=10dB) et classification obtenue d'après le modèle variationnel donné par (2.23).	64
3.1	Influence de la fonction d'arrêt $g(\nabla I)$ pour $g(t) = \frac{1}{1+t^2}$	69
3.2	Représentation du contour C par l'ensemble de niveau zéro d'une fonction u	70
3.3	Une partition de Ω	74
3.4	Approximations δ_α et H_α des distributions de Dirac et de Heaviside.	77
3.5	A chaque ensemble de régions Ω_i est associée une fonction de distance signée Φ_i . Les régions hachurées représentent les zones non nulles lorsque l'on applique une approximation de la distribution de Dirac δ_α ou de Heaviside H_α sur la fonction Φ_i	78
3.6	Etude locale de la vitesse du contour actif sur 2 régions. (a) Localement en a , $I(a) = \mu_2$ et la vitesse V est $(\mu_2 - \mu_1)^2 / \sigma_1^2$. (b) Localement en b , $I(b) = \mu_1$ et la vitesse V est $-(\mu_1 - \mu_2)^2 / \sigma_2^2$	80
3.7	Les deux types d'initialisation des ENZ : manuelle (à gauche) ou automatique (à droite).	86
3.8	Evolution des trois ENZ par courbure moyenne. Les paramètres λ et e_i de (3.37) sont nuls, $\gamma_i = 10.0 \forall i$, et $dt = 0.2$	90
3.9	Evolution des trois ENZ sans terme d'attache aux données ($e_i = 0.0 \forall i$).	90
3.10	Evolution des ENZ pour les 3 classes de moyennes $\mu_1 = 1.0$, $\mu_2 = 2.0$ et $\mu_3 = 3.0$	91
3.11	Evolution des régions associées à la figure 3.10.	92
3.12	Evolution des ENZ pour les 3 classes.	93
3.13	Evolution des régions associées à la figure 3.12.	94

3.14	Image "gdr" : évolution des ENZ pour les 3 classes.	95
3.15	Image "gdr" : évolution des régions associées à la figure 3.14.	96
3.16	Image "gdr" : évolution des trois ENZ (colonne de gauche) et des régions associées (colonne de droite) sur la version bruitée (RSB=3.2dB) des données synthétiques présentées sur la figure 3.14.	97
3.17	Influence de la fonction d'arrêt g sur la classification des données bruitées de la figure 3.16 en terme de pixels mal classés. Figure a : classification avec fonction d'arrêt g , figure b : classification sans fonction d'arrêt. Figure c : localisation des pixels mal classés lors de la classification sans fonction d'arrêt. Figure d : idem figure c pour une classification utilisant la fonction g . La figure e représente les pixels ayant obtenu une étiquette différente en utilisant ou pas g	98
3.18	Image IRM du cerveau contenant 4 classes. Nous montrons 4 étapes de l'évolution des ENZ (en haut) et des régions associées (en bas).	99
3.19	Données satellitaires SPOT (canal XS3) contenant 4 classes de paramètres pré-estimés. Nous montrons 3 étapes de l'évolution des ENZ. La dernière figure est le résultat de classification.	100
3.20	Données satellitaires SPOT (canal XS3) contenant 10 classes de paramètres pré-estimés.	101
3.21	Initialisations des ENZ pour les configurations A et B de la figure 3.22, et zoom sur la position des ENZ au niveau de la jonction triple.	104
3.22	Deux types de configurations initiales : A et B.	107
3.23	Influence du terme de restauration dans la classification par modèle actif.	108
4.1	Image "gdr" bruitée et vérité-terrain (VT) pour 3 et 8 classes, image 256x256 pixels.	119
4.2	Maximum de vraisemblance : résultats de classification pour 3 et 8 classes ainsi que nombre de pixels mal classés (#.).	120
4.3	Image "gdr" bruitée : résultats de classification pour 3 classes et 8 classes selon les différents modèles.	123
4.4	Nombre (#.) et localisation des pixels mal classés par les 4 modèles sur l'image "gdr" pour 3 et 8 classes. Le pourcentage de réussite est indiqué entre parenthèses.	124
4.5	Nombre (#.) et localisation des pixels différemment classés par les modèles M1 et M2 sur l'image "gdr" pour 3 et 8 classes.	125
4.6	Image "échiquier" (128×128 pixels) originale, et les trois versions bruitées.	125

4.7	Classification de l'image "échiquier" par MV (4 classes) pour différents niveaux de bruit.	125
4.8	Classification de l'image "échiquier" (4 classes) pour différents niveaux de bruit et selon les modèles. Nous indiquons le nombre de pixels mal classés ainsi que le pourcentage de réussite.	126
4.9	Résultats de classification pour l'image SPOT présentée sur la figure 2.17 du chapitre 2.	127
5.1	Principe de la classification automatique dans le cas multispectral.	134
5.2	Projection du potentiel vectoriel W sur le plan (I^1, I^2) dans le cas de 2 bandes et pour 3 classes.	138
5.3	Sélection de $i = \arg \min_{k=1, \dots, K} [\vec{f}(x) - \vec{\mu}_k]^T \Sigma_k^{-1} [\vec{f}(x) - \vec{\mu}_k]$ dans le cas où $P = 1$ et $K = 3$. Ici nous en déduisons $i = 1$ d'après (5.10). Le potentiel W final est la courbe représentée en pointillés.	140
5.4	Résultats de classification sur des données aériennes en couleur (fig. a) obtenus par l'algorithme de <i>fuzzy Cmeans</i> modifié (fig. b) et par l'algorithme présenté dans le tableau 5.2.	144
5.5	Résultats de classification sur des données aériennes en couleur (fig. a) obtenus par l'algorithme de <i>fuzzy Cmeans</i> modifié (fig. b) et par modèle de régions/contours actifs pour 2 types d'initialisation; init. 1 : initialisation des Φ_i dans des fenêtres 3×3 (fig. c); init.2 : pour des Φ_i initialisés par le résultat de classification donné par l'algorithme de <i>fuzzy Cmeans</i> modifié (fig. d).	150
5.6	Image couleur de l'IGN : évolution des régions (en haut) et de leurs contours (en bas) au cours des itérations pour une initialisation des Φ_i dans des fenêtres 3×3	151
5.7	Image couleur de l'IGN : évolution des régions (en haut) et de leurs contours (en bas) au cours des itérations pour une initialisation des Φ_i par le résultat de classification obtenu par <i>fuzzy Cmeans</i> modifié.	152
6.1	Structure de l'arbre tronqué.	156
6.2	Données multispectrales SPOT 2 de la baie de Lannion (Août 1997) après égalisation de l'histogramme.	161
6.3	Zones d'apprentissage (2719 pixels) dans lesquelles les paramètres des 8 classes ont été estimés et zones servant de validation (1466 pixels) pour les résultats de classification obtenus (vérité terrain).	162

6.4	Légende des thèmes d'occupation des sols et résultat de classification pour le modèle MV.	163
6.5	Résultats de classification pour les modèles ICM-N0, H-MAP, M1X et M2X.	164
B.1	Exemple de potentiel $\sum_{i=1}^K \varphi_c\left(\frac{(f(x)-\mu_i)^2}{2\sigma_i^2}\right)$ dans le cas de 5 classes.	178
B.2	Résultats de classification en fausses couleurs dans le cas de 9 classes (figure b) sur des données synthétiques bruitées à 10dB fournies par le GDR-PRC ISIS (figure a).	179
B.3	Résultats de classification en fausses couleurs dans le cas de 10 classes (figure b) sur des données SPOT fournies par le CNES (figure a).	180

Liste des tableaux

1.1	Caractéristiques des satellites SPOT 1, SPOT 2 et SPOT 3.	13
1.2	Algorithme du <i>K means</i> . L'indice n est relatif aux itérations effectuées. . . .	18
1.3	Echantillonneur de Metropolis.	25
1.4	Echantillonneur de Gibbs.	25
1.5	Algorithme du recuit simulé.	26
1.6	Algorithme ICM.	27
1.7	Estimateur du MPM.	27
2.1	Quelques fonctions φ proposées dans la littérature.	36
2.2	Algorithme associé à (2.12).	49
2.3	Paramètres (μ_i, σ_i) et couleur de l'étiquette pour chaque classe des images traitées.	55
3.1	Propriétés souhaitées sur la partition, solution du modèle de classification .	76
3.2	Algorithme associé au schéma (3.37).	85
4.1	Image "gdr" 8 classes : Pour chaque classe : nombre de pixels mal classés selon les modèles. La dernière ligne comptabilise le nombre de pixels d'échantillons par classe pour la vérité terrain (VT) de la figure 4.1.	121
4.2	Temps d'exécution et nombre d'itérations des différents modèles de classifica- tion pour l'image "gdr" avec 3 et 8 classes (temps cpu en secondes, processeur 300 Mhz). Ces 3 images ont une taille de 256x256 pixels.	122
5.1	Algorithme de <i>Fuzzy Cmeans</i> modifié.	135
5.2	Algorithme associé à la minimisation de (5.12) par rapport à (\vec{f}, b)	141
5.3	Algorithme associé au schéma (5.22).	148
6.1	Algorithme de minimisation de (6.1), extrait de [Chardin, 2000].	157
6.2	Nombre d'erreurs commises et pourcentages de réussite pour les différents modèles de classification	160

- 6.3 Pour chaque classe : nombre de pixels **mal** classés et pourcentage de **réus-**
site selon les modèles. La dernière ligne comptabilise le nombre de pixels
d'échantillons par classe pour la vérité terrain (VT). 167
- 6.4 Avantages et inconvénients des deux modèles de classification développés :
M1 est le modèle combinant restauration et classification, M2 est le modèle
dynamique de classification. 171

Notations et définitions

Généralités

- Les inconnues des modèles présentés au cours de ce travail sont des images définies sur un ensemble Ω que nous considérons inclus dans \mathbb{R}^n , $n = 2, 3$. La plupart du temps ces inconnues seront représentées par des fonctions à variables continues même s'il s'agit d'images qui, en pratique, renvoient à des notions discrètes.

D'un point de vue discret, le support Ω d'une image est modélisé par une grille de N pixels (ou *sites*) de taille nbl lignes par nbc colonnes, tel que $nbl * nbc = N$. $S = \{s\}_{s=1\dots N}$ est l'ensemble des sites ou pixels de l'image discrète, et $card(S) = N$. Les coordonnées d'un point sur la grille seront représentées par le couple (m, n) élément de $([1\dots nbl], [1\dots nbc])$.

- Dans ce qui suit, nous nous intéressons à la radiométrie des images. On représentera par la fonction continue $I : \Omega \rightarrow \mathbb{R}$, avec $I \in L^2(\Omega, \mathbb{R})$, l'intensité lumineuse associée à chaque pixel des données collectées. Ω , le support de l'image de frontière $\partial\Omega$, est supposé ouvert borné dans \mathbb{R}^n , avec $n = 2, 3$. La fonction I représente les données collectées (comme une image satellitaire par exemple).
- Dans un cadre continu, l'information contenue dans une image est modélisée par une fonction $f : \Omega \rightarrow \mathbb{R}$. Par défaut, $f \in L^\infty(\Omega, \mathbb{R})$. Dans le cas continu (respectivement discret), on notera $f(x)$ (resp. f_{mn}) la valeur de f au point $x \in \Omega$ (resp. la valeur de f au pixel de coordonnées (m, n)).
- Par abus de langage, et afin d'alléger les notations, nous confondrons parfois la fonction f avec sa valeur $f(x)$ comme par exemple dans l'expression des termes des fonctionnelles proposées où $\int_\Omega (f - I)^2 dx$ est utilisé en lieu et place de $\int_\Omega (f(x) - I(x))^2 dx$.

Définitions

- Id est la matrice identité.
- A^T , A^* et A^{-1} sont respectivement la transposée, l'adjointe et l'inverse de la matrice A .
- $\nabla = (\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n})^T$ est l'opérateur gradient.
- L'expression $f \sim \mathcal{N}(m, s)$ signifie que f suit une loi gaussienne de moyenne m et d'écart-type s .
- RSB est le Rapport Signal sur Bruit (SNR en Anglais) permettant de mesurer le rapport de variances de deux signaux s_1 et s_2 et défini par :

$$RSB = 10 \log_{10} \left[\frac{\sigma_a^2}{\sigma_{a-b}^2} \right],$$

avec σ_a^2 la variance du signal s_1 et σ_{a-b}^2 , la variance de la différence $s_1 - s_2$. Ce rapport sera utilisé par la suite pour caractériser le niveau de bruit d'une image dont la version originale non bruitée sera représentée par s_1 et la version bruitée par s_2 .

- $div(f) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}$ est l'opérateur de divergence appliqué à f .
- $\Delta f = \sum_{i=1}^n \frac{\partial^2 f}{\partial x_i^2}$ est l'opérateur Laplacien appliqué à f .
- $W^{1,p}(\Omega)$ est l'espace de Sobolev des fonctions dont la dérivée (au sens des distributions) est dans $L^p(\Omega)$.
On note $W^{1,2}(\Omega) = H^1(\Omega)$.
- La Variation Totale (TV) de f est définie par :

$$TV(f) = \int_{\Omega} |Df| = \sup \left\{ \int_{\Omega} f \cdot div(g); g \in C_0^1(\Omega, \mathbb{R}^n) \text{ et } |g(x)| \leq 1 \right\}.$$

si $f \in C^1(\Omega)$ alors $TV(f) = \int_{\Omega} |\nabla f|$, voir [Giusti, 1984] par exemple.

- $BV(\Omega)$ est l'espace des fonctions de Variation Totale Bornée (BV) [Giusti, 1984] :

$$BV(\Omega) = \{f \text{ telles que } TV(f) < +\infty\}.$$

Cet espace contient les fonctions pouvant être discontinues le long des courbes ($n = 2$), ce qui n'est pas le cas des espaces de Sobolev, et par conséquent $BV(\Omega)$ est l'espace adéquat pour l'étude des problèmes impliquant des ensembles de discontinuités tels que ceux liés à la segmentation d'images.

- Le périmètre de l'ensemble $A \subset \Omega$, au sens de De Giorgi [Giorgi, 1978], est défini par $Per_\Omega(A) = TV(\chi_A)$ où χ_A est la fonction caractéristique de A .
- $\mathcal{H}^{n-1}(\partial A)$ est la mesure $n - 1$ dimensionnelle de Hausdorff de l'ensemble ∂A , et si la frontière ∂A de A est lisse, nous avons :

$$\mathcal{H}^{n-1}(\partial A \cap \Omega) = Per_\Omega(A) = TV(\chi_A).$$

- **Γ -convergence** [Giorgi, 1978, Maso, 1992]

Définition : Soit X un espace métrique, $Y \subset X$.

Soit $g_\varepsilon : Y \longrightarrow [0; +\infty]$ une famille de fonctions dépendant de ε , $\varepsilon > 0$,

alors :

g_ε **Γ -converge** vers g quand $\varepsilon \rightarrow 0$ si :

$$\forall x \in X \begin{cases} \forall x_\varepsilon \rightarrow x, & \liminf_{\varepsilon \rightarrow 0} g_\varepsilon(x_\varepsilon) \geq g(x) \\ \exists x_\varepsilon \rightarrow x, & \limsup_{\varepsilon \rightarrow 0} g_\varepsilon(x_\varepsilon) \leq g(x). \end{cases}$$

Acronymes et abréviations

- EDP est l'acronyme d'"Equation aux Dérivées Partielles".
- ENZ est l'acronyme de "Ensembles de Niveaux Zéro" désignant l'ensemble :

$$\bigcup_{i=1}^K \{x \in \Omega / \Phi_i(x) = 0\},$$

pour Φ_i la fonction associée à la $i^{\text{ème}}$ classe.

- Le terme $p.p$ signifie "presque partout", i.e. partout sauf sur un ensemble de mesure de Lebesgue nulle.

Hypothèses et notations spécifiques à la classification

Nous définissons ici quelques notations utilisées ultérieurement en ce qui concerne la classification.

- L_s représente l'étiquette attribuée au site s .
- Λ est l'ensemble des étiquettes admissibles, $card(\Lambda) = K$ et $L_s \in \Lambda$.
- Lorsque cela n'est pas précisé, les indices i et j relatifs à la classification appartiennent à l'ensemble $[1, \dots, K]$.
- C_i est la $i^{ème}$ classe caractérisée par une étiquette dans Λ .
- Chaque pixel est caractérisé par son niveau d'intensité observé I_s ($I(x)$ en variable continue).
- La distribution est gaussienne à l'intérieur de chaque classe $C_{i;i=1..K}$, de moyenne μ_i et d'écart type σ_i . Dans le cas monospectral, nous utiliserons la notation vectorielle suivante :

$$\vec{\mu} = \{\mu_i\}_{i=1..K} \quad \text{et} \quad \vec{\sigma} = \{\sigma_i\}_{i=1..K}.$$

Dans le cas multispectral, où les données sont constituées de P bandes : $\vec{I} = [I^1, \dots, I^P]^T$, nous utiliserons les notations :

$$\vec{\mu}_i = \{\mu_i^p\}_{p=1..P} \quad \text{et} \quad \vec{\sigma}_i = \{\sigma_i^p\}_{p=1..P} \quad \text{pour} \quad i = 1, \dots, K.$$

- Le nombre K et les vecteurs de paramètres des classes sont connus ou estimés au préalable (i.e. classification *supervisée*).

Introduction

Le travail présenté ici s'inscrit dans le contexte général de l'aide au traitement et à l'interprétation des données. Les applications concernent plus particulièrement l'imagerie satellitaire. Cette thèse s'est déroulée au sein du projet ARIANA dont les activités de recherche concernent les problèmes inverses en observation de la Terre et cartographie. Parmi les motivations de ce projet, on trouve la volonté de comparer les méthodes issues de deux *écoles* : les approches probabilistes et les approches variationnelles déterministes.

Depuis quelques années, nous assistons à un véritable essor du marché des satellites civils avec une augmentation incessante des mises sur orbite. Le volume des données à traiter est par conséquent de plus en plus important et l'automatisation du traitement et de l'analyse de ces données devient incontournable. Les applications concernant l'imagerie satellitaire sont nombreuses (suivi de l'urbanisation, cartographie, gestion et aménagement des sols, systèmes d'information géographique SIG,...) et vont l'être encore plus dans les prochaines années avec l'arrivée sur le marché civil de satellites à haute et très haute résolution, donnant accès par exemple à l'étude et à l'analyse intra-urbaine.

La classification d'images est une opération d'autant plus importante qu'elle est présente à la base de bon nombre de chaînes d'analyse d'images. Elle consiste à découper les données en régions homogènes, chaque région étant repérée par une étiquette (numéro, couleur,...). Le travail réalisé au cours de cette thèse concerne le développement et la mise en œuvre de deux nouveaux modèles de classification à l'aide d'approches variationnelles. Le premier modèle s'inspire des travaux effectués sur la théorie de la Γ -convergence pour les problèmes impliquant des ensembles de discontinuités, alors que le second est fondé sur les travaux menés dans le cadre de la mécanique des fluides. Nous comparons ces deux modèles déterministes entre eux, mais également à des approches stochastiques développées au cours de ces dernières années.

Ce manuscrit est constitué de six chapitres organisés comme suit :

- Le premier chapitre est introductif et présente la problématique dans sa généralité. Nous commençons par définir la nature et les objectifs de l'opération de classification, ainsi que les notations utilisées au cours des chapitres suivants. Nous introduisons également les données satellitaires sur lesquelles nous appliquons les modèles proposés. Ce chapitre s'achève par une présentation de différents modèles de classification, en insistant plus particulièrement sur les approches stochastiques qui ont suscité de nombreux travaux dans ce domaine au cours de ces dernières années.
- Le second chapitre présente le premier modèle de classification par approche variationnelle que nous avons développé. Le chapitre débute par un rappel de différents travaux effectués dans le cadre des approches variationnelles pour la restauration et la reconstruction d'images au cours de ces dernières années. Nous introduisons ensuite la théorie de Van der Waals-Cahn-Hilliard élaborée en mécanique et sur laquelle repose en partie le modèle de classification que nous proposons. Le problème mécanique dont il est question est étudié d'un point de vue théorique par le biais de la Γ -convergence et nous exposons les principaux résultats publiés dans la littérature. Nous établissons le lien qui existe entre la stabilité d'un système mécanique constitué de fluides de différentes densités et une image composée de classes caractérisées par des distributions de luminosité différentes. Le modèle de classification que nous introduisons est défini à travers une suite de minimisation de fonctionnelles dont la solution finale est une partition régularisée des données. Lors de la convergence de cette suite de critères, le modèle change progressivement de comportement en commençant par restaurer l'image avant d'entamer le processus d'étiquetage des pixels. Nous terminons ce chapitre par une présentation des résultats obtenus sur des données synthétiques ainsi que sur des images satellitaires.
- Au sein du troisième chapitre, nous développons une autre approche variationnelle pour la classification d'images. Le modèle que nous définissons ici met en jeu un ensemble de régions et de contours actifs. Le critère proposé contient des termes liés à l'information sur les régions ainsi qu'à l'information au niveau des contours. Le chapitre débute par un rappel concernant les contours actifs et les travaux réalisés sur la formulation par ensembles de niveaux. Nous effectuons un état de l'art des récents travaux menés dans le cadre des approches dynamiques de type région, en insistant sur les avantages et inconvénients de chacune des méthodes. Nous détaillons ensuite les principales contraintes que nous souhaitons imposer sur la solution de partition, et nous traduisons ces contraintes en terme d'ensembles de niveaux. Nous aboutissons à une fonctionnelle définie sur une famille de

fonctions d'ensembles de niveaux. La minimisation de cette fonctionnelle conduit à la résolution d'un système d'équations aux dérivées partielles couplées et plongées dans un schéma dynamique. Il en résulte une évolution des régions actives convergeant vers une partition formée de classes homogènes séparées par des frontières régularisées. L'algorithme est testé sur un grand nombre de données synthétiques et réelles afin d'illustrer les propriétés de ce modèle.

- Le quatrième chapitre est consacré à l'étude comparative des résultats fournis par les modèles de classification proposés au sein des chapitres deux et trois. Nous comparons également nos résultats avec ceux fournis par deux modèles probabilistes dont l'optimisation est conduite à travers des algorithmes stochastiques. Le premier modèle stochastique contient un terme régularisant de type Potts ; le second, le chien-modèle, est plus complexe et permet de modéliser séparément les contours, les lignes et le bruit. Nous confrontons ainsi deux approches : variationnelle et probabiliste, ainsi que deux types d'algorithmes : déterministe et stochastique.

- L'extension au cas de données multispectrales des modèles présentés dans les chapitres 2 et 3 est l'objet du cinquième chapitre. Nous étendons ces modèles au cas de données constituées de P images, comme les trois bandes fournies par les satellites SPOT dans le mode XS. Nous insistons sur les différents termes concernés par cette extension pour les deux fonctionnelles proposées. Nous présentons également un algorithme nous permettant d'estimer au préalable le nombre ainsi que les paramètres des classes (l'algorithme du *Fuzzy Cmeans* avec contrainte d'entropie). Ainsi, en enchaînant un des deux modèles proposé à cet algorithme d'estimation, nous obtenons une méthode automatique de classification.

- Au sein du sixième chapitre, nous expérimentons les versions multispectrales des deux modèles que nous avons développés aux chapitres précédents sur une application réelle. Nous traitons une scène multispectrale SPOT 2 et nous comparons les résultats obtenus à ceux gracieusement fournis par le projet VISTA¹ de l'IRISA² Rennes. Une vérité terrain a également été mise à notre disposition, grâce à laquelle nous menons une étude détaillée des résultats obtenus par les différents modèles.

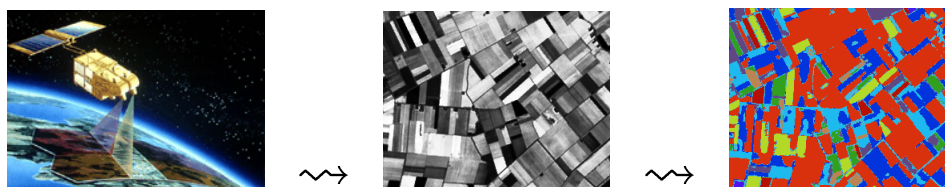
Ce manuscrit s'achève par une conclusion permettant de rappeler les principales contributions engendrées par ce travail, et de passer en revue les perspectives et extensions que nous envisageons dans l'avenir.

¹Vision Spatio-Temporelle et Active (<http://www.irisa.fr/vista>)

²Institut de Recherche en Informatique et Systèmes Aléatoires

Chapitre 1

Position du problème



© CNES

Nous présentons le contexte général dans lequel ce travail s'inscrit, ainsi que les différents problèmes abordés au cours de cette thèse. Nous exposons la problématique concernant la classification d'images en général en définissant la nature et les objectifs de l'opération de classification. Nous parlons ensuite brièvement du programme SPOT, initié il y a une vingtaine d'années, et des principales caractéristiques des données satellitaires traitées au sein de ce travail.

Nous terminons en passant en revue quelques modèles développés dans le cadre de la classification d'images au cours de ces dernières années. Une attention particulière est accordée à deux modèles stochastiques que nous utilisons dans le chapitre portant sur les résultats comparatifs.

1.1 La problématique

Ce paragraphe est consacré à la présentation générale de la classification d'image, mais aussi des données satellitaires qui sont traitées par la suite.

1.1.1 La classification d'images

La classification d'images est une opération présente à la base de nombreuses chaînes de traitement et d'interprétation de données. On distingue en général deux niveaux dans l'analyse d'images : les traitements de bas niveau et ceux de haut niveau. Dans les traitements de bas niveau, il s'agit d'extraire des attributs de l'image à partir des grandeurs physiques observées, que ces attributs soient réalistes ou non. Une fois ces attributs de bas niveau extraits, ils peuvent être interprétés par un traitement de haut niveau dans lequel est injecté de l'information ayant trait au monde réel. Par exemple, on pourra raisonnablement supposer à ce niveau qu'un aéroport se trouve très souvent à proximité d'une zone urbaine. Le processus de classification, dont il est question au cours de cette thèse, appartient aux approches de type bas niveau.

Nous nous intéresserons aux grandeurs photométriques des images traitées, et plus particulièrement à la valeur du niveau de gris retournée par chaque pixel (dans le cas de l'imagerie multispectrale, avec P bandes, nous aurons P niveaux de gris associés à chaque pixel). Cette grandeur physique est la plus simple, et la plus souvent utilisée, mais les modèles de classification que nous proposons peuvent s'étendre à des critères différents comme des attributs de texture par exemple (utilisation de filtres de Gabor [Jain et Farrokhnia, 1991] ou des matrices de cooccurrence [Connors *et al.*, 1984]), le tout étant que les attributs considérés puissent être caractérisés par des valeurs numériques.

Une fois ces attributs caractérisés, le principe de la classification consiste à réaliser une partition de l'image en régions repérées par des étiquettes différentes (numéro, couleur, ...).

Par convention, nous représenterons les différentes classes à l'aide de différentes couleurs.

Nous appellerons par la suite **partition régulière** un découpage de l'image en régions homogènes séparées par des frontières de longueur minimale. Cette contrainte sur la longueur des frontières rend l'opération de classification moins sensible aux petites régions ainsi qu'aux régions de frontière oscillante. Une classe peut être constituée de plusieurs régions non connexes, et nous distinguons par là, la classification de la segmentation qui consiste à découper l'image en régions connexes sans notion d'étiquetage.

Le processus de classification nécessite de déterminer le nombre de classes ainsi que les paramètres des classes caractérisant les différents attributs de l'image. Lorsque le nombre ainsi que les paramètres des classes sont connus au préalable (estimés par un algorithme ou donnés par un opérateur), nous parlerons de classification *supervisée*, au contraire de la classification *non supervisée* où le nombre de classes et leurs paramètres sont déterminés au sein même de la classification. Dans le travail qui suit, il s'agit de classification supervisée.

1.1.2 Présentation des données satellitaires

Les données satellitaires, sur lesquelles nous avons testé les modèles de classification proposés, proviennent des satellites lancés dans le cadre du programme SPOT. L'exploitation des images fournies par ces satellites concerne principalement la cartographie, l'inventaire des ressources naturelles et la gestion de l'occupation des sols en général. Les images nous ont été fournies dans le cadre de collaborations avec le CNES (Centre National d'Etudes Spatiales¹) ou achetées auprès de la société Spotimage².

Le programme SPOT (Satellite Pour l'Observation de la Terre) est réalisé par la France avec la participation de la Suède et de la Belgique. Ce programme a donné lieu au premier lancement du satellite SPOT 1 en 1986. Les deux satellites suivants, SPOT 2 et SPOT 3, ont respectivement été mis sur orbite en 1990 et 1993, et possèdent les mêmes caractéristiques que SPOT 1, celles présentées dans le tableau 1.1 (source : [CNES, 1989, Richards, 1993]). Concernant les satellites SPOT 1, SPOT 2 et SPOT 3, il existe deux modes spectraux : multispectral (mode *XS*) ou panchromatique (mode *P*). Le mode multispectral est caractérisé par 3 bandes spectrales, XS1, XS2 et XS3, soit 3 images de la même scène à des longueurs d'ondes différentes, dont la résolution est égale à 20 mètres. Le mode panchromatique ne retourne qu'une seule image de résolution égale à 10 mètres. La figure 1.1 montre la réflectance de quelques matériaux ainsi que la position des 3 bandes du mode XS

¹<http://www.cnes.fr>

²<http://www.spotimage.fr>

en fonction de la longueur d'onde. Par la suite, nous désignerons par "cas monospectral", le cas où nous ne traitons qu'une seule des 3 bandes du mode multispectral.

altitude		822 km	
cycle de passage		26 jours	
capteurs		2 HRV (Haute Résolution Visible) : - visée variable (système <i>pushbroom</i>) - visée oblique : $\pm 27^\circ$ (ou 460 km) - 2 modes : multispectral ou panchromatique	
mode multispectral		mode panchromatique	
nbre pixels/ligne	3000	nbre pixels/ligne	6000
fauchée	60 km	fauchée	60 km
résolution	20 m	résolution	10 m
Bandes	0.50 - 0.59 μm 0.61 - 0.68 μm 0.79 - 0.83 μm	Bande	0.51 - 0.73 μm

TAB. 1.1 – Caractéristiques des satellites SPOT 1, SPOT 2 et SPOT 3.

Le satellite SPOT 4 a été placé sur orbite en 1998 et possède une bande spectrale supplémentaire dans le moyen infrarouge. Le satellite SPOT 5, dont le lancement est prévu pour 2002, devrait offrir une meilleure résolution que ses prédécesseurs (2m50 et 5m en mode panchromatique et 10m en multispectral). A l'heure actuelle, seules des simulations d'images SPOT 5 sont disponibles, obtenues à partir d'un système aéroporté [Baudoin *et al.*, 1996, Pauc *et al.*, 1996]. Une résolution de 2m50 devrait offrir un certain nombre de possibilités concernant l'analyse urbaine, voire intra-urbaine (voir par exemple [Lorette, 1999]).

Notons que les modèles de classification que nous présentons ne sont pas contraints par la nature des données traitées. En effet, la seule hypothèse formulée sur les données concerne la nature du bruit : nous supposons que le bruit sur les données est additif, gaussien et de moyenne nulle. Nous ne prenons en compte aucun paramètre physique du satellite ou des capteurs. Il est cependant évident que cette hypothèse devra être modifiée dans le cas de l'imagerie radar où le bruit est multiplicatif et corrélé. En revanche, les algorithmes développés par la suite peuvent fonctionner sur divers types de données visible ou infrarouge fournies par différents satellites, comme LANDSAT par exemple.

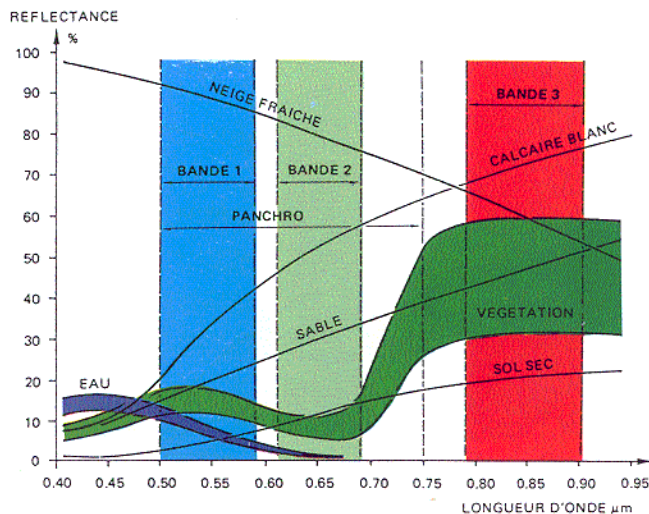


FIG. 1.1 – Réflectance spectrale de quelques matériaux et positionnement des 3 bandes XS1, XS2 et XS3 du capteur SPOT.

Nous avons également conduit, à titre indicatif, des expérimentations sur des données médicales IRM (Imagerie par Résonance Magnétique) pour montrer l'étendue du domaine d'application des modèles de classification que nous avons développés. Ces données nous ont été fournies par le Dr. Dormont de l'hôpital de La Pitié Salpêtrière.

1.2 Généralités sur les approches variationnelles et par EDP

Nous exposons ici quelques généralités concernant les approches variationnelles et par EDP, largement utilisées en traitement d'image. L'objectif de ce paragraphe n'est pas de faire un état de l'art, mais plutôt de donner un aperçu très général de la démarche utilisée par chacune des approches. Nous débutons les chapitres 2 et 3 par un état de l'art des travaux menés dans un cadre variationnel respectivement pour les problèmes impliquant un ensemble de discontinuités libres et pour la segmentation d'images par modèle dynamique.

L'ouvrage [Morel et Solimini, 1995] présente de nombreux travaux menés dans un cadre variationnel pour la segmentation d'images au cours de ces dernières années, et nous renvoyons le lecteur à [Deriche et Faugeras, 1996, Romeny, 1994, Weickert, 1998] pour une description détaillée d'un grand nombre d'approches par EDP utilisées dans les problèmes rencontrés en vision par ordinateur.

Comme son nom le laisse supposer, l'approche variationnelle fait appel à des outils mathématiques liés au calcul des variations ainsi qu'à l'analyse fonctionnelle. Le principe consiste à trouver la solution d'un problème par la minimisation d'une fonctionnelle. Cette fonctionnelle joue le rôle d'énergie, par analogie avec la physique, et contient différents termes qui doivent traduire les contraintes que l'on souhaite imposer sur la solution cherchée. Si le problème consiste à restaurer des données I corrompues par un bruit gaussien additif, nous pouvons définir la fonctionnelle :

$$J(f) = \int_{\Omega} \left(f(x) - I(x) \right)^2 dx + \int_{\Omega} |\nabla f(x)|^2 dx. \quad (1.1)$$

La minimisation de J doit conduire à une solution f proche des données I et dont les discontinuités, associées à de grandes valeurs de $|\nabla f(x)|$, sont minimisées. Cette contrainte régularisante est connue sous le nom de régularisation de Tikhonov [Tikhonov et Arsenin, 1977]. L'existence et l'unicité de la solution de minimisation fait appel à des théorèmes de l'optimisation des fonctionnelles, et nécessite de définir l'espace dans lequel on travaille. Par défaut, nous avons supposé que Ω est un ouvert borné de \mathbb{R}^n , $n = 2$ ou 3 , et pour la minimisation de (1.1) nous pouvons supposer que $f \in H^1(\Omega)$. Les conditions aux bords du support de l'image Ω doivent être choisies : symétriques, périodiques,... Nous supposons que $\frac{\partial f}{\partial \vec{n}} = 0$, $\forall x \in \partial\Omega$ avec \vec{n} la normale unitaire à la frontière $\partial\Omega$.

Le Lagrangien L est défini par :

$$J(f) = \int_{\Omega} L(f, \nabla f, x) dx. \quad (1.2)$$

Si f est solution de la minimisation de (1.1), et compte tenu des conditions aux bords choisies, alors f vérifie nécessairement l'équation d'*Euler-Lagrange* détaillée dans l'annexe A :

$$\nabla J = \frac{\partial L(f, \nabla f, x)}{\partial f} - \operatorname{div} \left(\frac{\partial L(f, \nabla f, x)}{\partial \nabla f} \right) = 0. \quad (1.3)$$

L'équation d'Euler-Lagrange est une condition nécessaire d'optimalité, et dans le cas de (1.1) elle donne :

$$\left(f(x) - I(x) \right) - \Delta f = 0 \quad (1.4)$$

La minimisation de (1.1) se ramène donc à la résolution de l'EDP (1.4) par le biais de l'équation d'Euler-Lagrange. Le lien entre minimisation de fonctionnelle et approche par EDP n'est pas toujours aussi évident. De nombreux travaux consistent à définir directement une EDP à résoudre, sans que celle-ci ne résulte forcément de la minimisation d'une fonctionnelle. C'est ce qui caractérise les *approches par EDP*. Nous y reviendrons plus en détails au début du chapitre suivant.

1.3 Quelques modèles de classification

Nous passons en revue un certain nombre de méthodes de classification utilisées dans diverses applications. Nous n'évoquons que des méthodes de classification supervisée ou tout au moins avec un nombre de classes connu au préalable. Le cinquième chapitre présente un algorithme d'estimation du nombre de classes et de leurs paramètres, étendu au cas multispectral.

Nous distinguons trois types d'approche : les méthodes fondées sur la minimisation de distance, les approches structurales, et les approches stochastiques. Nous ne présentons pas d'approche structurale, comme la croissance de régions, le lecteur intéressé pourra se référer aux ouvrages [Castleman, 1996, Cocquerez et Philipp, 1995, Gonzales et Woods, 1992, Jain, 1989] ou encore à [Pavlidis et Liow, 1988].

En introduction aux chapitres 2 et 3, nous faisons également un état de l'art concernant les approches variationnelles pour les problèmes à discontinuité libre (cf. chapitre 2) et pour les modèles dynamiques fondés sur des approches mixtes régions/contours (cf. chapitre 3).

1.3.1 Minimisation de distance

Le principe des méthodes reposant sur la minimisation de distance consiste à rechercher la classe la plus proche pour chaque pixel, ou groupe de pixels si l'on travaille dans une fenêtre d'analyse centrée sur le pixel courant. La notion de proximité est liée à la distance considérée. Ces méthodes sont très simples et souvent utilisées, mais ne sont pas robustes au bruit car elles ne comportent pas de terme de régularisation.

Compte tenu des hypothèses formulées précédemment, une classe est caractérisée par sa moyenne μ_i et son écart-type σ_i . Nous noterons respectivement m_s et e_s les moyennes et écart-types calculés dans la fenêtre d'analyse centrée sur le pixel courant s .

Méthodes non itératives

Ces approches consistent à parcourir l'image et à déterminer la classe la plus proche parmi les K possibles. En chaque site s de valeur x_s , nous calculons les K distances aux différentes classes. Cette distance peut tenir compte de la moyenne m_s mais aussi de l'écart-type σ_s calculés en ce site.

L'étiquette $k \in \Lambda$ associée à la classe C_k est attribuée au site s selon la formule :

$$k = \arg \min_{i=1 \dots K} d(x_s, C_i) \quad (1.5)$$

La classe C_i étant définie par sa moyenne μ_i et son écart-type σ_i , il existe plusieurs distances d possibles (voir [Basseville, 1985, Fukunaga, 1972, Kittler, 1975] par exemple) :

◦ distance euclidienne :

$$d_E(x_s, C_i) = (x_s - \mu_i)^2 \quad (1.6)$$

Cette distance ne permet pas de discriminer deux classes qui ne diffèrent que par leur écart-type.

◦ distance de Bhattacharyya ([Bhattacharyya, 1967, Kailath, 1967]) :

Elle permet de mesurer la distance entre deux lois de distribution :

$$d_B(x_s, C_i) = \frac{1}{4} \frac{(m_s - \mu_i)^2}{m_s + \sigma_s} + \frac{1}{2} \log \frac{m_s + \sigma_i}{2\sqrt{m_s \sigma_i}}. \quad (1.7)$$

◦ distance de Kullback ([Basseville, 1985]) dans le cas gaussien :

Comme la distance de Bhattacharyya, elle retourne la distance entre deux lois de probabilité

$$d_K(x_s, C_i) = \frac{1}{2} \frac{(m_s - \mu_i)^2}{m_s + \sigma_s} + \frac{1}{2} \left(\frac{m_s}{\mu_i} + \frac{\mu_i}{m_s} - 2 \right). \quad (1.8)$$

L'article [Basseville, 1985] passe en revue un grand nombre de distances permettant de déterminer en chaque pixel la classe la plus proche et d'attribuer ainsi l'étiquette correspondante. La classification par minimisation de distance est une technique très rudimentaire, mais utilisée dans certains systèmes d'information géographiques (SIG), pour leur simplicité et rapidité. En revanche, cette méthode n'est pas robuste au bruit que l'on rencontre sur bon nombre de données réelles.

Méthode itérative

Nous présentons ici un algorithme très utilisé et connu sous le nom de *K means* (ou *K moyennes*) [MacQueen, 1967]. Cette méthode est supervisée dans le sens où le nombre de classes doit être donné, mais pas nécessairement leurs paramètres. L'algorithme est donné dans le tableau 1.2. Lorsque les paramètres des classes ne sont pas connus, l'étape d'initialisation peut se résumer à sélectionner K sites de l'image pour initialiser les valeurs de μ_i^0 , et prendre $\sigma_i^0 = 1.0$, $\forall i$. Puis, on effectue des itérations jusqu'à convergence de l'algorithme. A chaque itération, on balaye l'image en attribuant une étiquette par minimisation

```

- Initialisation des paramètres  $\mu_i^0$  et  $\sigma_i^0$  pour  $i = 1 \dots K$ 
- Répéter  $n \leftarrow n - 1$  tant que le critère d'arrêt n'est pas vérifié
  en chaque site  $s$  calculer :
     $k_n = \arg \min_{i=1 \dots K} d(x_s, C_i^n)$ 
  pour  $i = 1 \dots K$  recalculer  $\mu_i^n$  et  $\sigma_i^n$  de chaque classe  $C_i^n$ 
  test sur le critère d'arrêt

```

TAB. 1.2 – Algorithme du *K means*. L'indice n est relatif aux itérations effectuées.

de distance. Il s'agit en général de la distance euclidienne, mais d'autres distances peuvent être choisies, notamment pour prendre en compte les écart-types des classes. Une fois les étiquettes attribuées, on calcule les nouvelles valeurs μ_i^n et σ_i^n pour les K classes trouvées à l'itération n :

$$\mu_i^n = \frac{1}{\text{card}(C_i^n)} \sum_{x_s \in C_i^n} x_s$$

$$\sigma_i^n = \frac{1}{\text{card}(C_i^n)} \sum_{x_s \in C_i^n} (x_s - \mu_i^n)^2.$$

On réitère ces opérations jusqu'à ce que le critère d'arrêt soit vérifié. Le critère d'arrêt peut porter sur le nombre relatif de sites dont l'étiquette a changé entre deux étapes, ou sur la variation relative des paramètres μ_i^n et σ_i^n entre les étapes n et $n + 1$. Comme pour les modèles précédents, ce modèle ne contient pas de terme de régularisation et l'algorithme est donc sensible à la présence de bruit.

1.3.2 Approches stochastiques

Nous présentons ici des approches définies dans un cadre stochastique. Dans ce type d'approche, les modèles sont souvent discrets et de nombreux travaux reposent sur la théorie des champs de Markov et l'estimation bayésienne [Berthod *et al.*, 1996, Bouman et Shapiro, 1994, Descombes *et al.*, 1997b, Descombes *et al.*, 1997a, Geman et Geman, 1984, Kato, 1994, Lakshmanan et Derin, 1989, Manjunath et Chellappa, 1991, Pérez, 1998]. Le lecteur intéressé par la théorie des champs de Markov et l'estimation bayésienne pourra se référer à [Chalmond, 2000, Cocquerez et Philipp, 1995, Descombes, 1993, Li, 1995].

Dans ce cadre, la configuration d'une image est vue comme une réalisation d'un champ aléatoire.

Champs de Markov, inférence bayésienne, règle de Bayes

Nous rappelons brièvement quelques résultats et définitions concernant les champs de Markov, ainsi que l'estimation bayésienne servant souvent de support à de nombreux modèles présentés dans un cadre stochastique. Les champs de Markov et l'estimation bayésienne sont largement utilisés en traitement des images. Les champs de Markov permettent de ramener des propriétés globales à des contraintes locales, et le théorème de Hammersley-Clifford en faisant l'équivalence entre champs de Gibbs et champs de Markov permet d'exprimer les contraintes locales sous forme de potentiel énergétique.

Soient S et Λ définis comme dans le chapitre notations. Soit X un champ aléatoire défini sur S , et à valeurs dans \mathbb{E}^N , où \mathbb{E} est l'espace des états (comme les niveaux de gris par exemple : $\mathbb{E} = [0, \dots, 255]$), et $N = \text{card}(S)$.

On associe à l'ensemble des sites S un *système de voisinage* V , qui, à chaque site s , associe V_s tel que :

- $s \notin V_s$
- $s \in V_t \Leftrightarrow t \in V_s$

X est un *champ de Markov* relativement à V si :

- $\forall x \in \mathbb{E}^{\text{card}(S)}, P(X = x) > 0$
- $\forall s \in S, P(X_s = x_s | X_t = x_t; t \in S \setminus \{s\}) = P(X_s = x_s | X_t = x_t; t \in V_s)$

Le *théorème de Hammersley-Clifford* [Besag, 1974, Geman et Geman, 1984] nous indique que si X est un champ de Markov relativement à V , alors sa distribution $P(X)$ est une distribution de Gibbs, soit :

$$P(X) = \frac{\exp(-U(X))}{Z}, \quad (1.9)$$

où Z est une constante de normalisation appelée "fonction de partition". U est un potentiel défini par :

$$U(X) = \sum_{q \in Q} U_q(X_s, s \in Q)$$

avec Q l'ensemble des cliques lié au système de voisinage V , et U_q un potentiel prenant en compte les interactions entre les sites d'une même clique, grâce auquel nous injectons de l'information a priori sur la solution cherchée X . Nous allons décrire deux modèles a priori : le modèle de Potts et le chien-modèle, .

Soit O une variable aléatoire représentant les données observées. Nous cherchons à estimer l'étiquetage discret de l'image représenté par le champ de Markov L relativement à un système de voisinage V . En chaque site s , nous devons attribuer une des K étiquettes $L_s \in \Lambda$ correspondant à la classe C_{L_s} de moyenne μ_{L_s} et d'écart-type σ_{L_s} . L'espace des états du champ L est Λ^N

Considérons la distribution de probabilité $P(L)$. Nous notons $R(L, L')$ la fonction de coût entre les deux configurations L et L' . L'estimation bayésienne consiste à minimiser le risque bayésien, c'est-à-dire à chercher la configuration minimisant l'espérance de la fonction de coût :

$$\hat{L} = \arg \min_{L'} \sum_{\Lambda^N} R(L, L') P(L/O) dL. \quad (1.10)$$

L'inconnue du problème étant L , pour les observations O connues, la loi *a posteriori* vérifie :

$$P(L/O) = \frac{P(O/L)P(L)}{P(O)} \quad \text{- règle de Bayes -} \quad (1.11)$$

la variable O étant donnée, nous avons :

$$P(L/O) \propto P(O/L)P(L) \quad (1.12)$$

$P(O/L)$ est appelé *vraisemblance* des données O , jouant le rôle d'attache aux données, et $P(L)$ est l'information a priori que l'on injecte sur l'étiquetage cherché (a priori d'homogénéité par exemple).

Nous rappelons deux critères fréquemment utilisés pour optimiser (1.10).

- Le MAP (Maximum A Posteriori) correspond à la fonction de coût

$$R(L, L') = 1 - \delta_{L'}(L), \quad (1.13)$$

où δ est le symbole de Kronecker. L'optimum est alors atteint pour la configuration qui maximise la loi a posteriori. L'algorithme stochastique le plus souvent utilisé pour le critère MAP est le recuit simulé [Laarhoven et Aarts, 1987] présenté par la suite.

- Le MPM (Modes Marginales a Posteriori) correspond à la fonction de coût

$$R(L, L') = \sum_{s \in S} 1 - \delta_{L'_s}(L_s). \quad (1.14)$$

L'optimum est alors atteint lorsque les différentes lois marginales en un site sont maximisées. Le critère MPM minimise en quelque sorte le nombre de sites mal classés [Marroquin *et al.*, 1987]. D'un point de vue algorithmique, le critère MPM conduit souvent à des algorithmes plus coûteux en temps calculatoires que dans le cas du MAP, mais les résultats

de classification sont souvent meilleurs, surtout lorsque le MAP est difficile à obtenir. Le principe algorithmique du MPM consiste à échantillonner la loi, et à ne conserver en chaque site que l'étiquette la plus représentée parmi les échantillons générés. Un algorithme d'optimisation selon un critère de type MPM est présenté par la suite.

Nous supposons que la vraisemblance $P(O/L)$ peut s'écrire comme le produit :

$$P(O/L) = \prod_{s \in S} P(O_s/L_s) \quad (\text{indépendance conditionnelle}).$$

Compte tenu des hypothèses gaussiennes formulées sur les classes, nous avons :

$$P(O_s/L_s) = \frac{1}{\sqrt{2\pi}\sigma_{L_s}} \exp\left[-\frac{(O_s - \mu_{L_s})^2}{2\sigma_{L_s}^2}\right]. \quad (1.15)$$

La distribution $P(L)$ dans (1.12) permet d'introduire un a priori sur le type d'interactions entre les cliques. Dans ce manuscrit, nous en décrivons deux : le modèle de Potts et le chien-modèle. Les propriétés markoviennes de L (cf. (1.9)) nous permettent d'écrire la loi a priori sur L de la façon suivante :

$$P(L) = \frac{1}{Z} \exp\left[-\sum_{q \in Q} U_q(L_s, s \in Q)\right]. \quad (1.16)$$

Modèle de Potts

La régularisation de Potts [Wu, 1982] consiste à imposer un a priori d'homogénéité $P(L)$ sur l'étiquetage cherché L . C'est une généralisation du modèle binaire d'Ising à des variables d'état m-aires (ici le nombre d'états m est $\text{card}(\Lambda)$).

Nous définissons :

$$P(L) = \exp\left[-\beta \sum_{\{s,r\} \in Q} \gamma(L_s, L_r)\right] \quad \text{- POTTs -}$$

$$\text{avec } \gamma(a, b) = \begin{cases} +1 & \text{si } a \neq b \\ 0 & \text{sinon} \end{cases}$$

Par conséquent, ce modèle a priori pénalise deux sites voisins possédant des étiquettes différentes, le poids de cette pénalité est de valeur fixée $\beta > 0$.

Dans les expérimentations menées par la suite avec une régularisation de type Potts, nous avons utilisé un modèle 4-connexe. De nombreuses synthèses de cet a priori ont été effectuées dans [Descombes, 1993], que ce soit en 4-connexité ou 8-connexité.

Cet a priori est largement utilisé dans le cadre des approches stochastiques, même s'il comporte certains inconvénients, notamment un phénomène de transition de phase [Georgii, 1988, Bellac, 1988, Morris *et al.*, 1996].

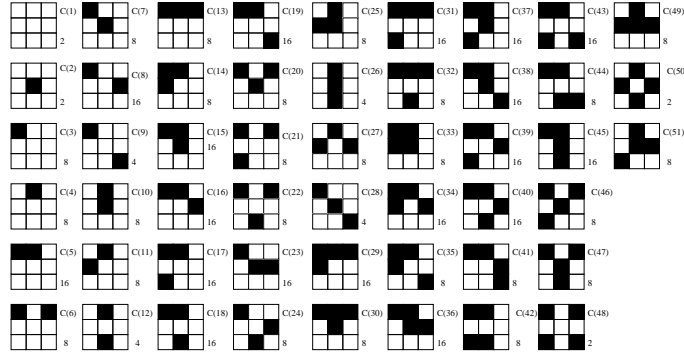


FIG. 1.2 – Les différentes configurations associées à un modèle binaire 3×3 et leur nombre d'éléments.

Le chien-modèle

Ce modèle a été proposé dans [Descombes *et al.*, 1995, Descombes *et al.*, 1997b, Descombes *et al.*, 1997a, Descombes *et al.*, 1999]. Son développement a été motivé, entre autres, par les limitations du modèle de Potts dues à la présence de transitions de phases largement étudiées en physique statistique [Georgii, 1988, Bellac, 1988, Morris *et al.*, 1996], mais également afin de mieux contrôler les contraintes imposées sur la solution.

Le principe consiste à considérer des cliques 3×3 pour pouvoir modéliser plus précisément les différentes configurations du voisinage (5×5) et représenter par exemple des structures comme les lignes et les contours. Des modèles définis sur de tels voisinages ont déjà été proposés, comme dans [Wolberg et Pavlidis, 1985].

Le chien-modèle est défini à partir de la discrimination entre le bruit, les contours et les lignes auxquels sont respectivement associés les paramètres n , e et l . Pour un système de cliques 3×3 et dans le cas d'un modèle binaire (deux états possibles), la figure 1.2 présente les 51 différentes classes de configurations $C(\cdot)$ et leurs nombres d'éléments, retenues parmi les $2^9 = 512$ possibles pour un ensemble de cliques 3×3 en tenant compte des symétries (inversion noir/blanc, rotation,...). A chacune des 51 classes est associé un paramètre définissant la valeur du potentiel des configurations de la classe. Un potentiel nul est imposé aux configurations uniformes ($C(1) = 0$). La construction du modèle consiste à imposer des contraintes en reliant les différents paramètres entre eux. Le premier type de contraintes concerne la pénalisation des contours. Par des considérations de symétries et de rotations, il suffit de définir trois orientations de contours, correspondant aux huit possibles induites par la taille des cliques. Les équations dérivées de ces contraintes sont représentées sur la figure 1.3. Ces contraintes sont définies pour des objets dont la largeur est d'au moins trois pixels. Lorsque l'objet considéré possède une largeur d'un pixel, il est assimilé à une ligne

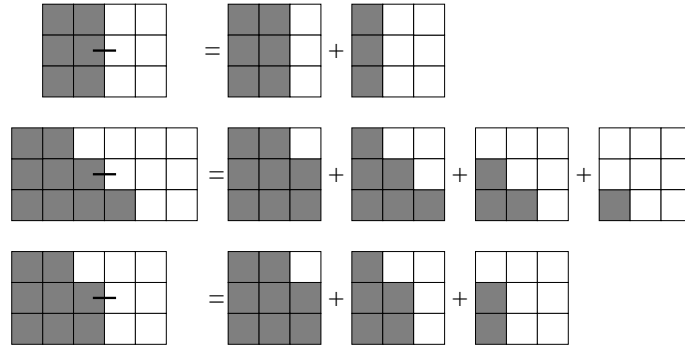


FIG. 1.3 – Equations liées aux contraintes sur les contours.

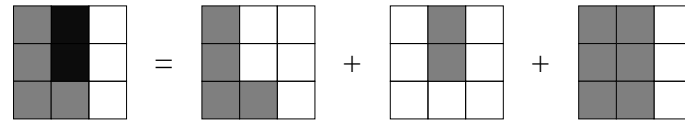


FIG. 1.4 – Extension m-aire du chien-modèle.

dont l'énergie associée est l . Pour des objets plus grands, (lignes doubles ou triples), l'énergie associée est $2e$. Toutes ces contraintes induisent onze équations dépendant de quatorze paramètres (voir [Descombes *et al.*, 1995] pour plus de détails) conduisant à la solution suivante :

$$\begin{aligned}
 C(3) = C(5) &= \frac{e}{4} & C(26) &= l - e & C(14) &= \frac{\sqrt{2}e}{4} \\
 C(16) = C(23) &= \frac{\sqrt{5}}{6}l - \frac{e}{4} & C(35) &= \frac{\sqrt{2}e}{2} \\
 C(11) = C(28) &= \frac{\sqrt{2}l}{3} - \frac{e}{6} & C(29) &= \frac{\sqrt{5}e}{6} \\
 C(13) = C(9) = C(19) &= \frac{e}{2}
 \end{aligned}$$

Le bruit est défini en attribuant l'énergie n à toutes les autres configurations.

L'extension de ce modèle au cas m-aire est réalisée en définissant l'énergie associée à une configuration comme une somme d'énergies données par le modèle binaire (voir l'illustration sur la figure 1.4). Le potentiel associé à chaque configuration est donc une combinaison linéaire des trois paramètres e , l et n :

$$\forall i = 0, \dots, 51 \quad C(i) = \epsilon(i)e + \lambda(i)l + \eta(i)n \quad \text{avec} \quad \epsilon(i), \lambda(i), \eta(i) \in \mathbb{R}. \quad (1.17)$$

La distribution (*a priori* sur la variable L) qui en résulte est donnée par :

$$P(L) = \frac{1}{Z(e, l, n)} \exp[-eN_0(L) - lN_1(L) - nN_2(L)] \quad \text{- CHIEN -} \quad (1.18)$$

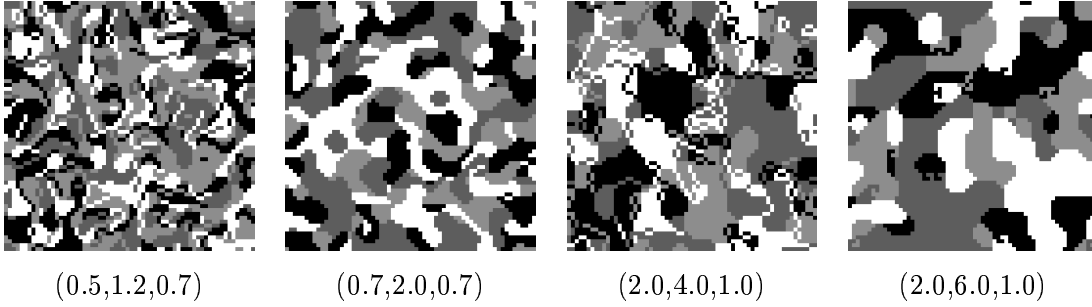


FIG. 1.5 – Réalisations du chien-modèle (4 classes) pour différentes valeurs du triplet (e, l, n) . Simulations réalisées par Xavier Descombes.

avec :

$$\begin{aligned}
 N_0(L) &= \sum_{i=1, \dots, 51} \epsilon(i) \#_i(L), \\
 N_1(L) &= \sum_{i=1, \dots, 51} \lambda(i) \#_i(L), \\
 N_2(L) &= \sum_{i=1, \dots, 51} \eta(i) \#_i(L).
 \end{aligned}$$

où $\#_i(L)$ est le nombre de configurations de type i dans la réalisation de L .

Le chien-modèle permet donc de contrôler en quelque sorte le taux de bruit, de contours et de lignes que l'on souhaite retrouver sur la solution de classification cherchée. Nous montrons sur la figure 1.5 quatre réalisations du chien-modèle pour différentes valeurs du jeu de paramètres (e, l, n) , obtenues par simulation. Cette figure révèle l'influence de chaque paramètre dans le contrôle des contours, des lignes et du bruit sur les réalisations obtenues. Ces trois paramètres devront être adaptés aux structures et à la morphologie des objets de l'image traitée, ce que ne permet pas de prendre en compte le modèle de Potts, l'a priori de Potts permettant seulement d'obtenir des régions plus ou moins homogènes en jouant sur la valeur du paramètre β introduit dans (1.17).

Optimisation et échantillonneurs

Nous présentons deux types d'échantillonneurs permettant de simuler le champs de Markov selon la loi a priori $P(L)$.

Echantillonneurs de Metropolis et de Gibbs

Les algorithmes associés à ces échantillonneurs sont représentés dans le tableau 1.3, pour

- Initialiser aléatoirement une configuration L^0 et fixer T (la température)
- Répéter $n \leftarrow n - 1$
 - balayer l'image de n'importe quelle manière, en chaque site s :
 - tirer aléatoirement une nouvelle étiquette l_s selon une loi uniforme dans Λ
 - calculer $dU = U_q(l_s) - U_q(L_s^{n-1})$
 - si $dU \leq 0$ alors $L_s^n = l_s$
 - sinon :
 - tirer ξ selon une loi uniforme dans $[0,1]$
 - si $\xi < \exp(\frac{dU}{T})$ alors $L_s^n = l_s$

TAB. 1.3 – Echantillonneur de Metropolis.

Metropolis [Metropolis *et al.*, 1953], et dans le tableau 1.4, pour Gibbs [Besag, 1974, Geman et Geman, 1984]. Dans ces deux tableaux, l'indice n est relatif à l'itération courante.

Une itération de l'échantillonneur de Gibbs est plus longue que celle de Metropolis du fait de la construction de la distribution conditionnelle locale pour les K éléments $i \in \Lambda$, mais l'échantillonneur de Gibbs nécessite moins d'itérations que celui de Metropolis pour converger. Pour les deux algorithmes d'échantillonnage, nous arrêtons les itérations pour un critère d'arrêt pouvant porter sur le nombre relatif de sites dont l'étiquette a été modifiée entre deux itérations successives n et $n + 1$.

Optimisation

- Initialiser aléatoirement une configuration L^0
- Répéter $n \leftarrow n - 1$
 - balayer l'image de n'importe quelle manière, en chaque site s :
 - calculer pour chaque élément i de Λ

$$p_i = \frac{1}{Z} \exp \left[- \sum_{q \in Q} U_q(L_s^n = i, s \in Q) \right]$$
 - tirer ξ selon une loi uniforme dans $[0,1]$
 - trouver j qui minimise l'expression :

$$\sum_{i=1}^j p_i > \xi \sum_{i=1}^K p_i$$
 - mettre à jour l'étiquette : $L_s^n = j$

TAB. 1.4 – Echantillonneur de Gibbs.

Nous présentons deux algorithmes d'optimisation de la loi a posteriori $P(L/O)$ définie par (1.12).

- Nous présentons dans un premier temps l'algorithme du recuit simulé [Laarhoven et Aarts, 1987] dans le cadre de l'estimateur du MAP. Cet algorithme stochastique est censé fournir une solution globale à l'optimisation de (1.10), au contraire des algorithmes déterministes d'optimisation comme par exemple la descente de gradient (gradient conjugué, gradient à pas fixe,...) ou comme l'ICM (*Iterated Conditional Mode* en Anglais) [Besag, 1974, Besag, 1986], dont le principe est présenté dans le tableau 1.6. Ces algorithmes déterministes n'atteignent généralement qu'une solution proche des conditions initiales choisies, cette solution pouvant être loin de la solution globale. Cette recherche d'un optimum global, lors d'un recuit simulé, se paie par un coût calculatoire et un temps d'exécution importants.

L'idée principale du recuit simulé consiste à faire dépendre la distribution de Gibbs $P(L/O)$ d'un paramètre T jouant le rôle de la température, et de diminuer lentement cette température pour *geler* les configurations et atteindre l'état d'équilibre du système. A chaque valeur de T , on écrit $P(L/O)$ comme une distribution de Boltzmann :

$$P(L/O) = \frac{1}{Z(T)} \exp - \frac{U(O, L)}{kT} \quad (1.19)$$

où k est la constante de Boltzmann. $U(O, L)$ est une énergie reliant les données O à l'étiquetage L (comme dans la relation (1.15) par exemple), et nous supposons ici que $U(O, L)$ contient également un a priori sur l'étiquetage L , comme le potentiel U_q défini dans (1.16). A chaque valeur de T nous échantillons L à l'aide d'un échantillonneur, comme ceux de Metropolis ou Gibbs par exemple. L'algorithme du recuit simulé permettant d'estimer l'étiquetage optimal \hat{L} , dans un cadre bayésien, est présenté dans le tableau 1.5. Les itérations

- 0 - Initialiser aléatoirement une configuration L^0
- 1 - Choisir une température initiale T_0
- 2 - Répéter $T^k \leftarrow T^{k-1}$
 - échantillonner L^k selon la loi a posteriori $P(L/O)$
- 3 - Faire décroître T^k et aller à l'étape 2

TAB. 1.5 – Algorithme du recuit simulé.

k de cet algorithme portent sur la décroissance de la température T . La loi de décroissance doit être relativement lente, la convergence vers l'optimum globale étant atteinte pour une

- Choisir une "bonne" configuration initiale L^0
- Répéter $k \leftarrow k - 1$ tant que le critère d'arrêt n'est pas vérifié
 - Balayer l'image, en chaque site s :
 - Si les sites voisins ont changé d'étiquette :
 - calculer l'énergie locale pour toutes les valeurs possibles des étiquettes L_s de Λ
 - choisir l'étiquette qui correspond à l'énergie minimale

TAB. 1.6 – Algorithme ICM.

température théoriquement nulle³. Le test d'arrêt peut porter sur le nombre relatif de sites dont l'étiquette a changé entre deux itérations.

- Dans le cadre de l'estimateur du MPM, nous présentons un schéma algorithmique dans le tableau 1.7. Le principe de cet algorithme consiste à générer une série de m échantillons de la variable L selon la loi a posteriori $P(L/O)$. On supprime ensuite les k premiers échantillons $L^0 \dots L^k$, puis nous attribuons l'étiquette la plus représentée parmi les $m - k$ échantillons en chaque site s . L'échantillonneur choisi pourra être celui de Gibbs ou de Metropolis par exemple. Supprimer les premiers échantillons permet d'éviter de prendre en considération des échantillons qui ne sont pas représentatifs de la loi de L .

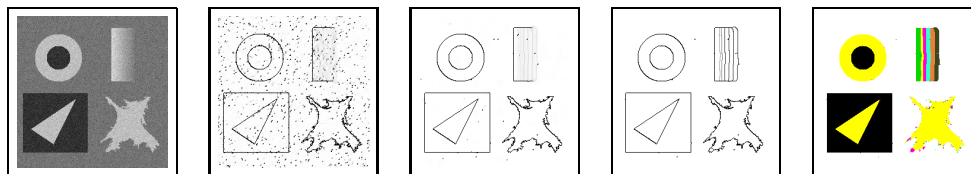
- 0 - Initialiser aléatoirement une configuration L^0
- 1 - Répéter $n \leftarrow n - 1$ jusqu'à $n = m$
 - échantillonner L^n selon la loi a posteriori $P(L/O)$
 - lorsque $n = m$, éliminer les k premiers échantillons L^0, \dots, L^k
 - pour chaque site s , attribuer l'étiquette la plus représentée dans les échantillons $(L_s^{k+1}, \dots, L_s^m)$

TAB. 1.7 – Estimateur du MPM.

³pour un refroidissement lent et une température initiale suffisamment élevée.

Chapitre 2

Classification avec restauration



Où nous présentons un modèle variationnel permettant de combiner un processus de classification à celui de restauration. Il en découle un modèle qui élimine progressivement le bruit sur une image tout en classant les pixels de celle-ci. Cette approche repose sur une suite de minimisation d'une famille de critères et fait appel à des notions de convergence pour les suites de fonctionnelles (Γ -convergence) introduites par De Giorgi en 1978.

Ce chapitre débute par un rappel des travaux effectués dans le cadre de la restauration d'image par approche variationnelle, avant de présenter la fonctionnelle combinant restauration et classification. Nous exposons quelques résultats mathématiques élaborés dans le domaine de la mécanique et offrant un cadre théorique pour l'étude du modèle que nous proposons. L'algorithme qui en découle est à la fois simple et efficace. Nous présentons un certain nombre de résultats de classification sur des images synthétiques ainsi que sur des données satellitaires.

2.1 Introduction

Nous présentons le premier modèle variationnel de classification que nous avons développé [Samson *et al.*, 2000b, Samson *et al.*, 1999d, Samson *et al.*, 1998, Samson *et al.*, 1999b]. Ce modèle permet de réaliser progressivement une restauration de l'image à l'aide d'un lissage anisotrope ainsi qu'un étiquetage des pixels. A titre indicatif, nous donnons en annexe B la première version de ce modèle, fondée sur l'intuition et donnant des résultats intéressants même s'il souffre d'un manque d'assise théorique.

2.2 Restauration par modèle variationnel

Nous rappelons ici quelques travaux effectués dans la cadre de la restauration d'images par approche variationnelle, en insistant plus particulièrement sur un modèle de lissage anisotrope proposé par Charbonnier *et al.* [Charbonnier *et al.*, 1997] dont nous nous inspirons pour la partie restauration dans le modèle que nous proposons.

Une grande partie des signaux collectés, en l'occurrence les images satellitaires, ne constituent qu'une version dégradée du signal original. Le signal émis par la source est soumis à deux principaux types de dégradation : l'influence de la réponse impulsionnelle du système physique d'acquisition du signal (ajout de flou sur les données, transformation des données...), et l'information parasite (ou *bruit*) venant se greffer sur l'information originale. La réponse impulsionnelle du système physique sera notée par un opérateur \mathcal{R} de $L(\Omega, \mathbb{R})$ l'ensemble des fonctions linéaires définies sur Ω et à valeurs réelles. Le bruit possède différentes natures (gaussien/impulsionnel,..., additif/multiplicatif,...) ainsi que différentes origines (capteur défectueux, sources parasites...). Le but de la restauration consiste à supprimer cette information parasite ainsi qu'à corriger l'effet de la réponse impulsionnelle du système, en évitant de dégrader l'information pertinente.

Nous formulons l'hypothèse suivante :

Hypothèse 1 :

une image est formée de zones homogènes
séparées par des ensembles de contours.

Il existe plusieurs approches pour modéliser la présence d'un contour, la plus simple étant de détecter un contour par la présence d'un contraste élevé de l'intensité. C'est le modèle de contour que nous utilisons dans ce chapitre. Il existe des modèles de contours plus élaborés dont un qui sera abordé plus en détails dans le chapitre suivant. Afin de modéliser le processus de dégradation de l'image source par du bruit, nous noterons par :

$$\begin{cases} f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R} & \text{l'image source,} \\ \nu : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R} & \text{le bruit,} \\ I : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R} & \text{les données bruitées observées.} \end{cases}$$

Le modèle de dégradation s'écrit (voir la figure 2.1)

$$I(x) = \mathcal{R}f(x) \odot \nu(x), \forall x \in \Omega \quad (2.1)$$

où \odot est un opérateur multiplicatif ou additif. Nous supposons par la suite que :

Hypothèse 2 :

$n = 2$, (imagerie bidimensionnelle)
 ν est un **bruit gaussien de moyenne nulle et additif**

Par conséquent, l'équation (2.1) devient :

$$I(x) = \mathcal{R}f(x) + \nu(x), \forall x \in \Omega \quad (2.2)$$

La restauration consiste à retrouver le signal original f à partir des observations I . L'opérateur \mathcal{R} est supposé connu, contrairement à la restauration *aveugle* où \mathcal{R} est inconnu [Chan et Wong, 1998, You et Kaveh, 1999]. Retrouver f à partir de l'équation (2.2) par la minimisation de la fonctionnelle

$$J_{\text{données}}(f, I) = \int_{\Omega} \left(\mathcal{R}f(x) - I(x) \right)^2 dx$$

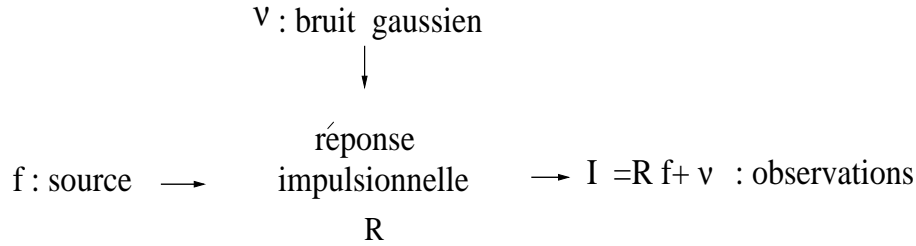


FIG. 2.1 – Modèle de dégradation.

est un problème mal posé au sens de Hadamard [Hadamard, 1923] d'où la nécessité d'introduire un terme de régularisation. Il s'agit d'imposer un *a priori* sur la solution recherchée afin de restreindre l'espace des solutions admissibles f . En l'occurrence, d'après l'hypothèse formulée ci-dessus, nous introduisons un *a priori* d'homogénéité sur la solution f . La fonctionnelle considérée s'écrit alors

$$J(f) = J_{\text{données}}(f, I) + \lambda^2 J_{\text{a priori}}(\nabla f)$$

le réel λ permet de jouer sur le poids du terme régularisant par rapport au terme d'attache aux données. Dans ce qui suit, la valeur de ce paramètre est déterminée empiriquement en fonction du poids que l'on souhaite accorder aux contraintes imposées sur la solution. Notons que l'on peut trouver une technique d'estimation de λ dans [Jalobeanu *et al.*, 1999]. Afin d'obtenir une solution lisse, le terme régularisant le plus simple est celui proposé par Thikonov [Tikhonov et Arsenin, 1977] : $J_{\text{a priori}}(f) = \int_{\Omega} |\nabla f(x)|^2 dx$. Cependant, cette contrainte pénalise toutes les discontinuités de la solution cherchée, dont les contours des objets. On parle alors de *lissage isotrope*. Afin d'atténuer cette contrainte sur les zones de contours, il convient d'introduire un terme régularisant dont l'effet est annihilé ou atténué sur les contours, conduisant à un *lissage anisotrope*. De nombreux auteurs ont travaillé sur de tels modèles comme par exemple dans [Aubert et Vese, 1997, Blake et Zisserman, 1987, Charbonnier *et al.*, 1997, Durand *et al.*, 1999, Kornprobst, 1998, March, 1992, March et Dozio, 1997, Morel et Solimini, 1995, Mumford et Shah, 1985, Teboul *et al.*, 1998, Lazaroia-Vese, 1996]. Le modèle que nous considérons est plus précisément développé dans [Charbonnier *et al.*, 1997] ainsi que dans [Aubert et Vese, 1997, Blake et Zisserman, 1987, Morel et Solimini, 1995], et la fonctionnelle que nous cherchons à minimiser est la suivante :

$$J(f) = \int_{\Omega} (\mathcal{R}f(x) - I(x))^2 dx + \lambda^2 \int_{\Omega} \varphi(|\nabla f(x)|) dx. \quad (2.3)$$

La fonction $\varphi : \mathbb{R} \rightarrow \mathbb{R}^+$ doit avoir au moins les deux propriétés suivantes (voir la figure 2.2 ainsi que le tableau 2.1) : lorsque $t \in \mathbb{R}$ est proche de zéro, $\varphi(t)$ est quadratique (lissage

isotrope), et pour de grandes valeurs de t , $\varphi(t)$ est sous-quadratique. De plus, on suppose que $\forall t \in \mathbb{R}$, $\varphi(t) \geq 0$; $\varphi(0) = 0$; $\varphi(t) = \varphi(-t)$; φ est continue et différentiable et $\forall t \geq 0$, $\varphi'(t) \geq 0$.

Si f est un minimum de J , nécessairement f vérifie l'équation d'Euler-Lagrange associée à (2.3) :

$$\mathcal{R}^*(\mathcal{R}f(x) - I(x)) + \lambda^2 \operatorname{div} \left(\frac{\varphi'(|\nabla f(x)|)}{2|\nabla f(x)|} \nabla f \right) = 0, \quad \forall x \in \Omega. \quad (2.4)$$

On retrouve dans l'expression de cette EDP, le rôle joué par la fonction de régularisation φ à travers le coefficient de diffusion $\frac{\varphi'(|\nabla f(x)|)}{2|\nabla f(x)|}$ dans le terme de divergence. Ainsi, la diffusion est pondérée par un terme dont la valeur varie en fonction de la présence ou non d'un contour. Sur les zones homogènes, le coefficient de diffusion est constant, donnant lieu à un lissage isotrope, et sur les zones de contours, la diffusion est atténuée en fonction de la valeur du module du gradient.

Notons que plusieurs travaux ont été menés dans le cadre des approches par EDP pour la restauration et la segmentation d'images. Le principe consiste à définir directement le modèle à travers la résolution d'une EDP de type diffusion anisotrope comme (2.4), sans que celle-ci ne résulte de la minimisation d'une fonctionnelle comme c'est le cas pour (2.4). Nous ne détaillons pas ces travaux, mais nous renvoyons le lecteur intéressé par les approches par EDP à [Alvarez *et al.*, 1992, Catté *et al.*, 1992, Deriche et Faugeras, 1996, Nitzberg et Shiotani, 1992, Nordström, 1990, Perona et Malik, 1990, Romeny, 1994, Weickert, 1998, You *et al.*, 1996]

On classe généralement les fonctions φ en fonction de leur convexité (voir le tableau 2.1). Pour les fonctions convexes, l'étude théorique de la minimisation de la fonctionnelle J conduit à des résultats tels que l'existence d'un minimum (voir par exemple [Lazaroaia-Vese, 1996] pour $\varphi(t) = \sqrt{1+t^2}$ et [Chambolle et Lions, 1997] pour $\varphi(t) = |t|$). En revanche dans le cas non convexe, l'étude théorique des solutions de minimisation de J est bien plus difficile, même si les meilleurs résultats sont obtenus pour de telles fonctions.

Hypothèse 3 :

Nous supposons par la suite que $\mathcal{R} = Id$, c'est-à-dire que nous ne prenons pas en compte l'effet de la réponse impulsionnelle du système physique traversé par la source (sauf au paragraphe 2.7)

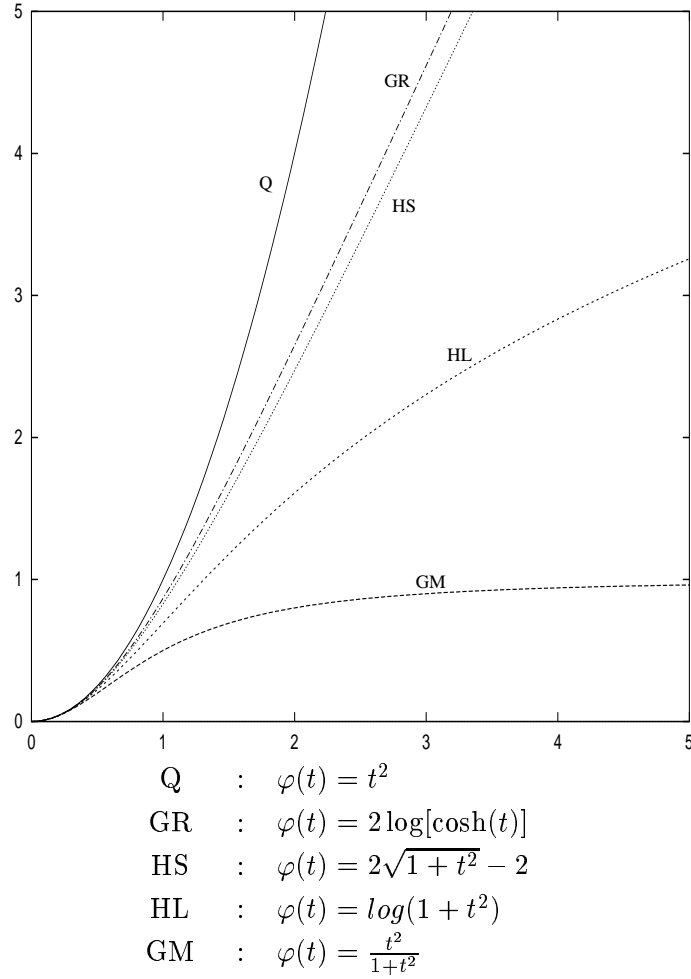


FIG. 2.2 – Exemple de fonctions de régularisation (voir le tableau 2.1).

Dans ce contexte, Mumford et Shah [Mumford et Shah, 1985] ont introduit une fonctionnelle qui porte leur nom et qui reste une référence en ce qui concerne les problèmes de segmentation par approche variationnelle. Son étude a suscité et suscite toujours un grand intérêt tant au niveau applicatif qu'au niveau théorique [Ambrosio et Tortorelli, 1990, Ambrosio et Tortorelli, 1992, Aubert et Vese, 1997, Braides, 1998, March, 1992, March et Dozio, 1997, Morel et Solimini, 1995, Teboul *et al.*, 1998]. En plus de la notion de restauration, on trouve dans le modèle qu'ils proposent la notion de segmentation, c'est-à-dire obtenir une partition de l'image. L'expression de la fonctionnelle de Mumford-Shah est la suivante :

$$J_{MS}(f, B) = \int_{\Omega} \left(f(x) - I(x) \right)^2 dx + \lambda^2 \int_{\Omega \setminus B} |\nabla f(x)|^2 dx + \mathcal{H}^{n-1}(B), \quad f \in BV(\Omega) \quad (2.5)$$

	$\varphi(t)$	$\frac{\varphi'(t)}{2t}$	convexité
Tikhonov [Tikhonov et Arsenin, 1977]	t^2	1	oui
Total Variation [Rudin <i>et al.</i> , 1992]	$ t $	$\frac{1}{2 t }$ (si $t \neq 0$)	oui
Geman & McClure [Geman et McClure, 1985]	$\frac{t^2}{1+t^2}$	$\frac{1}{(1+t^2)^2}$	non
Hebert & Leahy [Hebert et Leahy, 1989]	$\log(1 + t^2)$	$\frac{1}{1+t^2}$	non
Green [Green, 1990]	$\log(\cosh(t))$	$\frac{\tanh(t)}{2t}$ (si $t \neq 0$)	oui
Hyper Surfaces [Charbonnier <i>et al.</i> , 1997]	$2\sqrt{1+t^2} - 2$	$\frac{1}{\sqrt{1+t^2}}$	oui
Perona & Malik [Perona et Malik, 1990]	$1 - \exp(-t^2)$	$\exp(-t^2)$	non

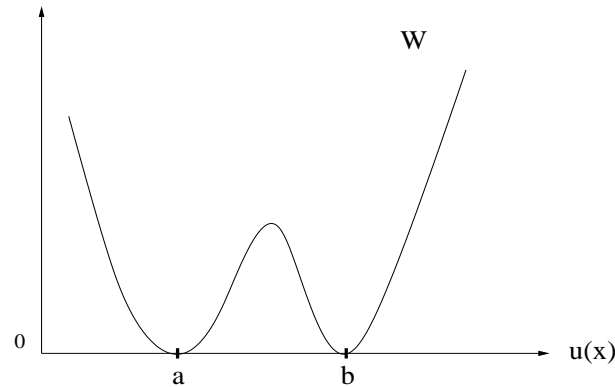
TAB. 2.1 – Quelques fonctions φ proposées dans la littérature.

avec B désignant l'ensemble des discontinuités de f . Le couple solution (f, B) de la minimisation de J_{MS} est une image f composée de zones homogènes séparées par un ensemble de discontinuités B de longueur minimale. La minimisation est rendue difficile par le fait que l'ensemble B est inconnu dès le départ. Cette fonctionnelle entre dans le cadre plus général des problèmes impliquant des ensembles de discontinuités libres (*free discontinuity problems*). Des théorèmes fondés sur la Γ -convergence permettent d'approcher la fonctionnelle J_{MS} par une suite de fonctionnelles elliptiques, comme par exemple dans [Ambrosio et Tortorelli, 1990, Ambrosio et Tortorelli, 1992, Braides, 1998, Gobbino, 1998, March, 1992, March et Dozio, 1997, Teboul *et al.*, 1998]. La minimisation du critère J_{MS} conduit à une partition de l'image formée de zones homogènes séparées par des contours de longueur minimale.

Le modèle que nous proposons dans ce chapitre doit permettre de réaliser simultanément la restauration ainsi que la classification de l'image. Par conséquent, en plus des notions de partition et de restauration, nous souhaitons introduire la notion d'étiquetage de l'image, c'est-à-dire attribuer une étiquette à chaque pixel. Pour ce faire, nous proposons d'ajouter un terme à la fonctionnelle J définie dans (2.3) qui joue le rôle de contrainte de classification. L'expression générale de la fonctionnelle est la suivante :

$$J(f) = \int_{\Omega} \left(f(x) - I(x) \right)^2 dx + \lambda^2 \int_{\Omega} \varphi(|\nabla f(x)|) dx + \eta^2 J_{\text{classif.}}(f, \Theta), \quad (2.6)$$

où Θ est un jeu de paramètres caractérisant les différentes classes, en l'occurrence compte tenu des hypothèses formulées précédemment, Θ contient les moyennes et les variances des

FIG. 2.3 – Exemple de potentiel W à deux puits.

K classes. Ainsi, $\Theta = (\vec{\mu}, \vec{\sigma})$, et nous rappelons que dans le cas de K classes, $\vec{\mu} = [\mu_1, \dots, \mu_K]$ et $\vec{\sigma} = [\sigma_1, \dots, \sigma_K]$. Le réel η permet de jouer sur le poids accordé au terme $J_{\text{classif.}}(f, \Theta)$. Ce terme doit permettre de contraindre l'ensemble des valeurs $f(x)$ à prendre une des K valeurs des étiquettes associées aux classes. La valeur des étiquettes étant la moyenne correspondante de chaque classe, la solution f de la minimisation de (2.6) sera composée de régions homogènes de niveaux $\{\mu_i\}_{i=1\dots K}$ séparées par des bords francs.

2.3 La théorie de Van der Waals-Cahn-Hilliard

La théorie de Van der Waals-Cahn-Hilliard sur les transitions de phases a été largement étudiée en mécanique pour décrire les états stables des systèmes physiques constitués de phases instables. Par souci de clarté, nous présentons le cas de systèmes constitués de deux phases mais la théorie reste valable pour les systèmes contenant K phases différentes (voir par exemple [Baldo, 1990]).

Le système physique composé des deux phases instables peut modéliser par exemple un alliage chimique tel que *Fer-Aluminium* [Allen et Cahn, 1979], ou encore un fluide dont la densité peut prendre deux valeurs [Modica, 1987]. La problématique commune à ces différents modèles consiste à caractériser les solutions de stabilité de ces systèmes, et plus particulièrement le comportement des interfaces entre les deux composants du système étudié.

2.3.1 Présentation et principaux résultats

Considérons un système physique constitué d'un fluide dont l'énergie libre de Gibbs par unité de volume est une fonction notée W dépendant de la distribution de densité u du

fluide. Si le fluide est constitué de deux phases différentes décrites par les niveaux $u(x) = a$ et $u(x) = b$, alors le potentiel W possède deux minima. W est un potentiel à deux puits (*double-well potential* en Anglais), comme l'illustre la figure 2.3. Le fluide est supposé être dans des conditions isothermes et contenu dans un domaine borné Ω . La masse du fluide est supposée être constante. A stabilité, le fluide ne prendra que les deux valeurs $u(x) = a$ et $u(x) = b$. L'approche consistant à caractériser l'état stable du système par l'étude de $\inf_u \int_{\Omega} W(u(x))dx$ sous la contrainte $\int_{\Omega} u(x)dx = m$, est confrontée au problème de la multiplicité des solutions u de minimisation. La théorie de Van der Waals-Cahn-Hilliard aboutit à l'introduction d'un terme de perturbation (ou de *régularisation*) $\varepsilon|\nabla u|^2$, avec ε petit, permettant de restreindre l'espace des solutions admissibles u . Il s'agit d'introduire un a priori pondéré d'homogénéité sur les solutions possibles u .

On considère alors la suite de problèmes variationnels P_{ε} suivante (cf. [Baldo, 1990, Bellettini *et al.*, 1991, Fonseca et Tartar, 1989, Modica, 1987, Sternberg, 1991, Sternberg et Zeimer, 1994]) :

$$P_{\varepsilon} \begin{cases} \inf_{u \in H^1(\Omega)} E_{\varepsilon}(u), \\ E_{\varepsilon}(u) = \int_{\Omega} \left[\varepsilon |\nabla u(x)|^2 + \frac{1}{\varepsilon} W(u(x)) \right] dx, \\ \text{avec la contrainte suivante sur } u : \quad \int_{\Omega} u(x)dx = m, \end{cases} \quad (2.7)$$

où Ω est un ouvert borné de \mathbb{R}^n de frontière lipschitzienne continue, et m est la masse totale du fluide. $|\nabla u(x)|$ représente le module du gradient de la distribution de densité u . Le potentiel $W : \mathbb{R} \rightarrow \mathbb{R}^+$ est une fonction continue positive possédant exactement deux minima de valeurs a et b ($a < b$) tels que $W(a) = W(b) = 0$. De plus, W est quadratique autour de chaque minimum et croît au moins linéairement à l'infini¹. Notons que cette étude a également été développée pour le cas $W : \mathbb{R}^P \rightarrow \mathbb{R}^+$, avec $P > 1$, comme dans les travaux [Baldo, 1990, Fonseca et Tartar, 1989, Sternberg, 1991]).

Concernant l'étude du problème P_{ε} et de l'ensemble des solutions minimisant la famille de fonctionnelles E_{ε} lorsque $\varepsilon \rightarrow 0^+$, les auteurs de [Bellettini *et al.*, 1991, Fonseca et Tartar, 1989, Modica, 1987, Sternberg, 1991] ont utilisé la théorie de la Γ -convergence développée par De Giorgi [Giorgi, 1978, Maso, 1992] pour l'étude de suites de fonctionnelles. Nous rappelons la définition de la Γ -convergence :

Définition :

Soit X un espace métrique, $Y \subset X$.

Soit $g_{\varepsilon} : Y \rightarrow [0; +\infty]$ une famille de fonctions dépendant de ε , $\varepsilon > 0$,

alors :

¹ voir [Fonseca et Tartar, 1989] par exemple pour l'expression explicite des conditions sur W

g_ε Γ -converge vers g quand $\varepsilon \rightarrow 0$ si :

$$\forall x \in X \begin{cases} \forall x_\varepsilon \rightarrow x, & \liminf_{\varepsilon \rightarrow 0} g_\varepsilon(x_\varepsilon) \geq g(x) \\ \exists x_\varepsilon \rightarrow x, & \limsup_{\varepsilon \rightarrow 0} g_\varepsilon(x_\varepsilon) \leq g(x). \end{cases}$$

Cette convergence variationnelle a été largement utilisée en traitement d'images, notamment pour approcher le terme de longueur de l'ensemble des discontinuités B contenu dans la fonctionnelle (2.5), définie par Mumford et Shah dans la cadre de la segmentation, par une suite de fonctionnelles elliptiques, comme par exemple dans [Ambrosio et Tortorelli, 1990, Ambrosio et Tortorelli, 1992, Braides, 1998, Gobbino, 1998, March, 1992, March et Dozio, 1997, Teboul *et al.*, 1998].

Si Ω et W vérifient les conditions exposées précédemment, les principaux résultats fournis par ces auteurs de [Bellettini *et al.*, 1991, Fonseca et Tartar, 1989, Modica, 1987, Sternberg, 1991] sont les suivants :

- E_ε Γ -converge vers E_0 avec :

$$E_0(u) = \begin{cases} \rho \text{Per}_\Omega(R_1) & \text{si } u \in BV(\Omega) \text{ et } u(x) \in \{a; b\} \text{ p.p.} \\ +\infty & \text{sinon,} \end{cases}$$

où $R_1 = \{x \in \Omega / u(x) = a\}$ et ρ est défini par :

$$\rho = 2 \inf_g \left\{ \int_{-1}^1 \sqrt{W(g(s))} |g'(s)| ds; \ g \text{ est } C^1 \text{ par morceaux et } g(-1) = a, g(1) = b \right\}.$$

- si u_ε est une suite de solutions de P_ε telles que u_ε converge vers u_0 dans $L^1(\Omega)$, alors u_0 est solution du problème :

$$\inf_u \text{Per}_\Omega(R_1) \quad \text{avec} \quad \begin{cases} u \in BV(\Omega), \\ W(u(x)) = 0 \text{ p.p.}, \\ \int_\Omega u(x) dx = m \end{cases}$$

- Toute suite (v_ε) telle que $E_\varepsilon(v_\varepsilon) \leq \text{constante} < \infty \ \forall \varepsilon$, admet une sous-suite qui converge dans $L^1(\Omega)$ (compacité dans $L^1(\Omega)$).

Le potentiel W peut être considéré comme une contrainte de niveaux forçant la fonction u à prendre exclusivement une des deux valeurs $u_\varepsilon(x) = a$ ou $u_\varepsilon(x) = b$. Le poids de cette contrainte de niveaux augmente lorsque ε tend vers zéro par valeurs positives. Le terme de

perturbation $\varepsilon|\nabla u(x)|^2$ induit un effet régularisant sur la solution u_ε en évitant la formation de singularités sur u_ε . Cette contrainte de régularité diminue lorsque ε tend vers zéro. Nous rappelons qu'il existe des résultats similaires pour le cas de potentiels W possédant K puits et pour des systèmes composés de P phases instables (i.e. $W : \mathbb{R}^P \rightarrow \mathbb{R}^+$), le lecteur intéressé pourra consulter par exemple les travaux de [Angenent et Gurtin, 1989, Baldo, 1990, Bronsard et Reitich, 1993, Reitich et Soner, 1996, Sternberg et Zeimer, 1994].

Remarque : L'EDP dynamique associée à la minimisation de E_ε , à savoir l'équation dynamique donnée par $\frac{\partial v}{\partial t} = -\nabla E_\varepsilon(v)$, est l'équation de réaction-diffusion de Ginzburg-Landau sur $v : (x, t) \in \Omega \times \mathbb{R}^+ \rightarrow v(x, t) \in \mathbb{R}$ définie par :

$$\begin{cases} \frac{\partial v}{\partial t} = 2\varepsilon\Delta v - \frac{1}{\varepsilon}W'(v), & \text{où } t \text{ représente le paramètre de temps,} \\ v(x, 0) = v_0(x). \end{cases} \quad (2.8)$$

Avec v_0 la version initiale de v au temps $t = 0$. De nombreux auteurs se sont penchés sur l'étude de cette équation, indépendamment du problème de minimisation (2.7), comme dans [Barles *et al.*, 1992, Caginalp, 1989, Rubinstein *et al.*, 1989]. Dans ces travaux, il est montré que la solution $u_\varepsilon(x, t)$ de (2.8) quand $\varepsilon \rightarrow 0$ est une fonction constante par morceaux ne prenant que les deux valeurs a et b . De plus, l'évolution de l'interface entre les deux régions $\{x \in \Omega / u_\varepsilon(x, t) = a\}$ et $\{x \in \Omega / u_\varepsilon(x, t) = b\}$ suit une évolution par courbure moyenne, ou *mean curvature motion* [Allen et Cahn, 1979, Barles *et al.*, 1992, Kichenassamy *et al.*, 1996, Reitich et Soner, 1996, Rubinstein *et al.*, 1989], ce qui induit une certaine régularité de l'interface.

2.3.2 Analogies avec la classification d'images

Nous allons utiliser les résultats exposés dans le cadre de la mécanique pour la classification et la restauration d'une image. Dans ce contexte, nous cherchons à caractériser la répartition d'intensité lumineuse représentée par la fonction $f : \Omega \rightarrow \mathbb{R}$ par analogie avec l'étude de la distribution de densité u dans (2.7).

Supposons que l'image soit constituée de deux classes représentées par les régions

$$\begin{aligned} R_1 &= \{x \in \Omega / f(x) \sim \mathcal{N}(a, \sigma_a)\} \\ R_2 &= \{x \in \Omega / f(x) \sim \mathcal{N}(b, \sigma_b)\}, \end{aligned}$$

i.e. sur R_1 l'intensité est distribuée de façon gaussienne autour de la moyenne a , et autour de b sur R_2 . Les régions R_1 et R_2 peuvent être constituées d'ensemble non nécessairement

connexes. Compte tenu de l'équation (2.2), des notations utilisées, et d'après (2.7), nous considérons le problème suivant :

$$\tilde{P}_\varepsilon \begin{cases} \min_f \tilde{E}_\varepsilon(f), \\ \tilde{E}_\varepsilon(f) = \int_\Omega \left[\varepsilon |\nabla f(x)|^2 + \frac{1}{\varepsilon} W(f(x)) \right] dx, \\ \text{avec la contrainte sur } f : \int_\Omega (f(x) - I(x))^2 dx \leq \sigma_\nu^2 \end{cases} \quad (2.9)$$

où σ_ν est l'écart-type du bruit additif gaussien de moyenne nulle ν de (2.2). W est un potentiel vérifiant les propriétés exposées précédemment et prenant en compte la répartition gaussienne de l'intensité dans chaque classe. La contrainte impose à la solution f_ε de rester en moyenne proche des données radiométriques I , "proche" au sens de la distance quadratique.

Compte tenu des résultats précédents, lorsque $\varepsilon \rightarrow 0^+$, la solution f_ε du problème \tilde{P}_ε converge vers une image composée de zones homogènes dont les pixels sont éléments de R_1 ou R_2 , séparées par des frontières régulières (i.e. de longueurs minimales) et minces (plus ε est petit, plus minces sont les frontières). La minimisation de la longueur des contours rappelle le terme portant sur la longueur des discontinuités B définis par Mumford et Shah dans (2.5) pour la segmentation d'images.

Lorsque ε est fixé, la fonctionnelle de \tilde{P}_ε est constituée de deux termes :

$$\tilde{E}_\varepsilon(f) = \int_\Omega \left[\underbrace{\varepsilon |\nabla f|^2}_{\text{terme de régularisation}} + \underbrace{\frac{1}{\varepsilon} W(f)}_{\text{contrainte de niveaux}} \right] dx. \quad (2.10)$$

Le premier terme de \tilde{E}_ε pénalise l'apparition de discontinuités sur la solution. Dans la cadre de la restauration d'images, ce terme régularisant a été initialement introduit par Tikhonov [Tikhonov et Arsenin, 1977], et induit un lissage isotrope sur l'image. Par ailleurs, le second terme de la fonctionnelle restreint l'ensemble des valeurs admissibles de l'intensité $f_\varepsilon(x)$ au couple de valeurs $\{a; b\}$, moyennes des deux classes. Ainsi, même si le poids accordé au terme régularisant est important, le potentiel W permet de maintenir les niveaux de la solution proche des valeurs caractérisant les deux classes.

La convergence sur ε induit une évolution du poids relatif des deux termes. L'influence de la contrainte de niveaux évolue inversement à celle du terme de régularisation. La valeur initiale de ε est suffisamment importante² pour rendre le terme de contrainte de niveaux négligeable par rapport à celui de régularisation. Au cours de la décroissance de ε , l'influence du potentiel W augmente progressivement en même temps que diminue celle de

²"importante" dans le sens "non négligeable"

la régularisation. On peut dire grossièrement qu'on assiste à un changement progressif du comportement du modèle au cours de la convergence sur ε ; initialement, le modèle est de type restauration plutôt que de type classification, et lors de la décroissance de ε la tendance s'inverse progressivement. Schématiquement, l'étiquetage des pixels de l'image n'a lieu qu'une fois l'image suffisamment restaurée. Selon le niveau de bruit de l'image, il s'avèrera utile de laisser plus de temps à la phase de restauration et donc d'opter pour une décroissance lente du paramètre ε .

De plus, d'après les travaux menés sur l'étude de l'équation de réaction-diffusion (2.8), l'interface entre les régions R_1 et R_2 suit une évolution par courbure moyenne, traduisant ainsi la régularisation géométrique opérée sur l'ensemble des contours.

La contrainte imposée sur f dans (2.9) mesure l'adéquation entre les observations I et la solution cherchée. Par conséquent, la solution f_ε du problème \tilde{P}_ε de (2.9) lorsque ε tend vers zéro par valeurs positives, converge vers une image, tenant compte bien évidemment des données, composée des régions homogènes R_1 et R_2 séparées par des interfaces de longueurs minimales.

Ces résultats restent valables dans le cas de K classes et donc pour des potentiels W formés de K puits.

2.4 Modèle proposé

2.4.1 Expression de la fonctionnelle

Compte tenu de ce qui précède, nous définissons une famille de fonctionnelles en introduisant un potentiel W ainsi que le paramètre de convergence ε dans (2.6) :

$$J_\varepsilon(f) = \underbrace{\int_{\Omega} \left(f(x) - I(x) \right)^2 dx}_{\text{attache aux données}} + \varepsilon \lambda^2 \underbrace{\int_{\Omega} \varphi(|\nabla f(x)|) dx}_{\text{terme de restauration}} + \frac{\eta^2}{\varepsilon} \underbrace{\int_{\Omega} W(f(x); \vec{\mu}, \vec{\sigma}) dx}_{\text{contrainte de niveaux}}, \quad (2.11)$$

et le problème associé consiste à trouver \tilde{f} vérifiant :

$$\tilde{f} = \lim_{\varepsilon \rightarrow 0^+} \left[\arg \min_f J_\varepsilon(f) \right]. \quad (2.12)$$

Lorsque ε est fixé, les deux premiers termes de (2.11) sont ceux du modèle de restauration avec préservation des discontinuités semblable à la fonctionnelle (2.3). Le dernier terme est une contrainte de niveaux inspiré des travaux menés dans le cadre de la théorie de Van der Waals-Cahn-Hilliard sur les transitions de phases. Le potentiel W restreint l'espace des valeurs prises par $f_\varepsilon(x)$ à celui des K étiquettes des classes en tenant compte du caractère

gaussien de la distribution d'intensité sur chaque classe. D'après les conventions choisies (cf. paragraphe 1.1 du chapitre 1), W attire les valeurs de $f_\varepsilon(x)$ vers celles des étiquettes $\{\mu_1, \dots, \mu_K\} = \Lambda$.

Le potentiel W défini dans (2.11) doit prendre en compte les paramètres gaussiens de la distribution de l'intensité dans les classes, et doit vérifier les conditions exposées dans le paragraphe 2.3.

W possède K minima situés en $f_\varepsilon(x) = \mu_i$ pour $i = 1, \dots, K$, et s'annule en ces points. D'après les hypothèses présentées dans le paragraphe 2.3 page 38, nous avons construit W comme une fonction parabolique par morceaux et telle que $W \in C^1(\mathbb{R})$. Comme l'illustre la figure 2.4, autour de μ_i l'expression de W est de la forme $W(y) = P_i(y) = \frac{w_i}{\sigma_i^2}(y - \mu_i)^2$, $y \in \mathbb{R}$, où w_i est une constante positive et avec μ_i, σ_i supposés connus. Puis, nous construisons la parabole $Q_i(y) = k_i - c_i(y - p_i)^2$ joignant P_i et P_{i+1} respectivement aux points α_i et β_i (Q_i est une *jonction parabolique*). En écrivant les conditions pour obtenir une jonction de classe $C^1(\mathbb{R})$, nous aboutissons au système :

$$\begin{cases} P_i(\alpha_i) = Q_i(\alpha_i) \\ P_{i+1}(\beta_i) = Q_i(\beta_i) \\ P_i'(\alpha_i) = Q_i'(\alpha_i) \\ P_{i+1}'(\beta_i) = Q_i'(\beta_i), \end{cases}$$

avec $P_i' = \frac{\partial P_i}{\partial y}$ et $Q_i' = \frac{\partial Q_i}{\partial y}$. Nous obtenons ainsi un système constitué de 4 équations et 5 inconnues : $\alpha_i, \beta_i, k_i, c_i$ ($c_i > 0$) et p_i , soit un degré de liberté : nous avons choisi de rendre l'intervalle $[\alpha_i, \beta_i]$ aussi petit que possible pour donner plus d'importance aux paraboles P_i . Nous obtenons un potentiel qui respecte les conditions énoncées dans le paragraphe 2.3, à savoir : W est quadratique autour de chaque minimum, $W(\mu_i) = 0$, $\forall i = 1, \dots, K$, et W croît plus que linéairement à l'infini. Dans le cas de K classes, l'expression de W est la suivante :

$$W(y) = \begin{cases} \frac{w_1}{\sigma_1^2}(y - \mu_1)^2 & \text{si } y \leq \alpha_1 \\ \frac{w_K}{\sigma_K^2}(y - \mu_K)^2 & \text{si } y \geq \beta_{K-1} \\ \frac{w_i}{\sigma_i^2}(y - \mu_i)^2 & \text{si } \beta_{i-1} \leq y \leq \alpha_i \quad i = 1, \dots, (K-1) \\ k_i - c_i(y - p_i)^2 & \text{si } \alpha_i \leq y \leq \beta_i \quad i = 1, \dots, (K-1) \end{cases} \quad (2.13)$$

Notons que dans le cas d'un mélange de classes, il est possible d'ajuster manuellement la valeur du paramètre w_i afin d'obtenir des puits bien marqués. Plus la valeur de w_i est importante, plus le puits est étroit.

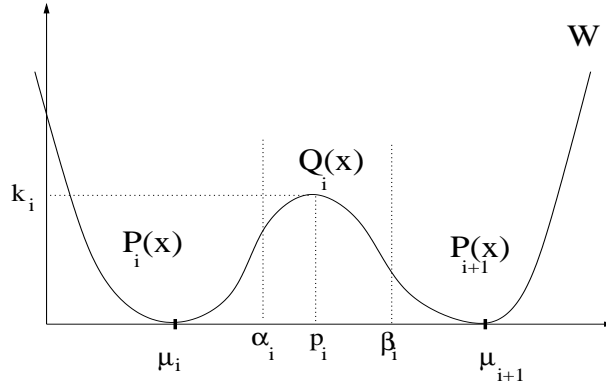


FIG. 2.4 – Construction du potentiel quadratique par morceaux W .

Pour chaque jeu de données $(\vec{\mu}, \vec{\sigma})$, nous obtenons par construction un potentiel différent. La construction de W est effectuée avant l'exécution de l'algorithme de minimisation associé au problème (2.12).

On suppose que la fonction φ de (2.11) est une fonction régularisante (cf tableau 2.1) possédant les caractéristiques exposées dans le paragraphe 2.2.

Soit ε fixé, si f_ε minimise J_ε alors f_ε est une image restaurée formée de zones homogènes avec des bords francs (à condition que φ ne soit sous-quadratique à l'infini), et sur chaque région les pixels ont une valeur proche d'une des K étiquettes μ_i , $i = 1, \dots, K$. Lors de la décroissance de ε , la minimisation de J_ε induit dans un premier temps une restauration des données qui s'estompe progressivement pour laisser émerger le processus d'étiquetage.

Remarque : la classification est obtenue comme solution limite d'une suite de minimisation de fonctionnelles par rapport à une fonction à variable continue (comme la fonction d'intensité f_ε), et non discrète comme c'est souvent le cas dans les modèles stochastiques de classification [Berthod *et al.*, 1996, Descombes *et al.*, 1997b, Descombes *et al.*, 1997a, Kato, 1994].

2.4.2 Justification

Par la suite, lorsque cela ne porte pas à confusion, nous omettrons l'indice ε de f_ε , même si la solution de minimisation f dépend de ε .

Nous allons examiner la consistance théorique du modèle proposé en se référant aux résultats établis dans le cadre de la théorie des transitions de phases. Les travaux concernant

les modèles de type Van der Waals-Cahn-Hilliard concernent le problème (2.7), alors que nous proposons le problème (2.12). Nous avons donc effectué quelques extrapolations par rapport à ces travaux que nous allons examiner ici, en distinguant celles qui sont théoriquement justifiées de celles qui ne le sont pas.

• **Du terme de contrainte au terme de pénalité :**

La contrainte $\int_{\Omega} u(x)dx = m$ dans (2.7) est introduite directement dans la fonctionnelle (2.11) sous forme de terme de pénalité (terme d'attache aux données).

Dans [Baldo, 1990, Bellettini *et al.*, 1991, Modica, 1987, Sternberg, 1991] les résultats théoriques présentés sur l'étude du problème (2.7) et des solutions limites associées, sont obtenus par le biais de la théorie de Γ -convergence. Une des propriétés intéressantes de la Γ -convergence, réside dans le fait que les résultats de convergence restent valables si on ajoute une perturbation continue (par rapport à la topologie considérée) à la suite de fonctionnelles. En particulier, dans le cas de la topologie (forte) de L^2 on a :

lorsque $\varepsilon \rightarrow 0^+$:

$$\begin{array}{ll} \text{si} & F_{\varepsilon}(f) \xrightarrow{\Gamma} F_0(f) \\ \text{alors} & F_{\varepsilon}(f) + \int_{\Omega} \left(f(x) - I(x) \right)^2 dx \xrightarrow{\Gamma} F_0(f) + \int_{\Omega} \left(f(x) - I(x) \right)^2 dx \end{array}$$

où $I : \Omega \rightarrow \mathbb{R}$ est une fonction de $L^2(\Omega)$. De plus, les éléments de la suite de minima f_{ε} de E_{ε} dans (2.7), ont les caractéristiques exposées dans le paragraphe 2.3 et restent proches des données I au sens de la distance quadratique.

• **D'un potentiel possédant 2 puits à un potentiel possédant K puits :**

Le potentiel défini dans (2.11) possède K puits centrés sur les moyennes $\mu_i; i = 1, \dots, K$. La plupart des auteurs qui se sont penchés sur les modèles de type Cahn-Hilliard ont étudié le cas de potentiels à 2 puits [Allen et Cahn, 1979, Barles *et al.*, 1992, Caginalp, 1989, Modica, 1987, Rubinstein *et al.*, 1989, Sternberg, 1991]. Quelques uns ont travaillé sur le cas spécifique de 3 phases comme dans [Reitich et Soner, 1996, Sternberg et Zeimer, 1994], et l'extension au cas général de potentiels à K puits a été réalisée par S. Baldo dans [Baldo, 1990]. Il a étendu les résultats montrés dans le cas de 2 phases au cas où la distribution de densité dans (2.7) est $u : \mathbb{R}^n \rightarrow \mathbb{R}^P$ et lorsque le potentiel est défini par $W : \mathbb{R}^P \rightarrow \mathbb{R}^+$ avec K minima de valeurs nulles. Lorsque $u : \Omega \rightarrow \mathbb{R}$ et pour un potentiel W possédant K

minima, l'expression de E_0 , la Γ -limite de E_ε définie dans (2.7) est :

$$E_0(u) = \sum_{i,j=1}^K d(\mu_i, \mu_j) \mathcal{H}^1(\Gamma_{ij}) \begin{cases} \text{si } E_0(u) < +\infty, \\ u \in BV(\Omega), \\ W(u(x)) = 0 \text{ p.p.} \end{cases} \quad (2.14)$$

avec Γ_{ij} l'interface entre les régions $\Omega_i = \{x \in \Omega / x \in C_i\}$ et $\Omega_j = \{x \in \Omega / x \in C_j\}$. La distance d est définie par :

$$d(\mu_i, \mu_j) = \inf_g \left\{ \int_0^1 \sqrt{W(g(s))} |g'(s)| ds \begin{cases} g \in C^1([0; 1], \mathbb{R}), \\ g > 0, g(0) = \mu_i, g(1) = \mu_j. \end{cases} \right\} \quad (2.15)$$

Ainsi, le problème limite de minimisation de E_ε est un problème de minimisation de la longueur pondérée de l'ensemble des interfaces entre les régions associées à chaque classe. Si u_ε converge dans $L^1(\Omega; \mathbb{R})$ vers u_0 lorsque $\varepsilon \rightarrow 0^+$, alors :

$$u_0(x) = \sum_{i=1}^K \mu_i \chi_{\Omega_i}(x) \quad (2.16)$$

avec χ_{Ω_i} la fonction caractéristique de l'ensemble Ω_i .

Nous reviendrons dans le chapitre 5 sur l'expression de cette Γ -limite dans le cas vectoriel $u : \Omega \rightarrow \mathbb{R}^P$ pour l'imagerie multispectrale (P bandes).

• D'une perturbation quadratique à la régularisation par fonction φ :

Le terme de perturbation $\varepsilon |\nabla u|^2$ dans (2.7) est remplacé par le terme de régularisation de type $\varepsilon \varphi(|\nabla f|)$ dans (2.11).

Remplaçons $\varepsilon |\nabla u|^2$ dans (2.7) par $\Theta : \Omega \times \mathbb{R} \times \mathbb{R}^n \rightarrow \Theta(x, u, \varepsilon \nabla u) \in \mathbb{R}^+$. Owen et Sternberg [Owen et Sternberg, 1991] ont prouvé que si Θ est convexe par rapport à $\varepsilon \nabla u$ (il existe d'autres conditions sur Θ dans leur article), et si Θ possède un comportement au moins quadratique par rapport à $\varepsilon \nabla u$ à l'infini, alors u_ε la solution de P_ε lorsque $\varepsilon \rightarrow 0$ possède les mêmes propriétés que pour une perturbation isotrope $\varepsilon^2 |\nabla u|^2$ (i.e. la perturbation de (2.7) modulo un changement d'échelle sur ε). Ainsi, pour les fonctions régularisantes du tableau 2.1 autres que celle de Tikhonov, l'étude des solutions de (2.12) reste un problème ouvert.

2.5 Minimisation et algorithme

Si f minimise 2.11, alors nécessairement f annule le gradient de la fonctionnelle J_ε , et nous obtenons l'équation d'Euler-Lagrange :

$$\begin{cases} \left(f(x) - I(x) \right) + \frac{\eta^2}{2\varepsilon} W'(f(x)) - \varepsilon \lambda^2 \operatorname{div} \left(\frac{\varphi'(|\nabla f(x)|)}{2|\nabla f(x)|} \nabla f(x) \right) = 0, & \forall x \in \Omega, \\ \frac{\partial f(x)}{\partial n} = 0, & \forall x \in \partial\Omega, \end{cases} \quad (2.17)$$

où n est la normale externe au domaine $\partial\Omega$. Notons la non linéarité de l'EDP (2.17) car $W'(f(x))$ est linéaire par morceaux et lorsque φ n'est pas la fonction quadratique, le terme de divergence n'est pas linéaire.

Afin d'obtenir un lissage avec préservation des contours, nous imposons les conditions suivantes sur la fonction régularisante φ (cf. [Charbonnier *et al.*, 1997]) :

- (a) $\lim_{t \rightarrow 0} \frac{\varphi'(t)}{2t} = \gamma > 0$ et $\gamma < \infty$,
- (b) $\lim_{t \rightarrow \pm\infty} \frac{\varphi'(t)}{2t} = 0$,
- (c) $\frac{\varphi'(t)}{2t}$ symétrique strictement décroissante sur $[0; +\infty[$.

La condition (a) induit un lissage isotrope sur les zones homogènes, alors que la condition (b) permet d'arrêter ou d'atténuer le processus de lissage en présence de contours. La condition (c) est destinée à éviter les instabilités dans le schéma algorithmique.

Supposons ε fixé. L'approche que nous adoptons pour minimiser J_ε par rapport à f repose sur la méthode de *régularisation semi-quadratique* [Charbonnier *et al.*, 1997, Geman et Reynolds, 1992]. Cette méthode permet de surmonter la difficulté liée à la non linéarité introduite par le terme de divergence de l'EDP (2.17). L'idée principale consiste à introduire une variable auxiliaire, liée à la présence de contours, qui permet de se ramener à un système d'EDP semi-linéaires. Soit φ une fonction régularisante vérifiant les conditions (a), (b) et (c) exposées plus haut. Dans [Charbonnier *et al.*, 1997, Geman et Reynolds, 1992], il est montré que φ peut se mettre sous la forme :

$$\varphi(t) = \inf_b [bt^2 + \psi(b)] \quad (2.18)$$

avec ψ une fonction dont l'expression est déduite de φ : $\psi(b) = g((g')^{-1}(b)) - b(g')^{-1}(b)$ où $g(t) = \varphi(\sqrt{t})$, et ψ est strictement convexe³.

$b : \mathbb{R}^2 \rightarrow \mathbb{R}$ est une variable auxiliaire dont la valeur réalisant l'infimum de (2.18) est :

$$b = \frac{\varphi'(t)}{2t}. \quad (2.19)$$

³voir [Charbonnier *et al.*, 1997, Geman et Reynolds, 1992] pour plus de détails

On peut alors écrire la fonctionnelle $J_\varepsilon(f)$ sous la forme :

$$J_\varepsilon(f) = \min_b J_\varepsilon^*(f, b) \quad (2.20)$$

avec

$$J_\varepsilon^*(f, b) = \int_{\Omega} \left(f(x) - I(x) \right)^2 dx + \varepsilon \lambda^2 \int_{\Omega} \left[b(x) |\nabla f(x)|^2 + \psi(b(x)) \right] dx + \frac{\eta^2}{\varepsilon} W(f(x)) dx. \quad (2.21)$$

$J_\varepsilon^*(f, b)$ est appelée *énergie augmentée*. L'intégrande de $J_\varepsilon^*(f, b)$ est convexe par rapport à la variable auxiliaire b (pour f fixé), et lorsque b est fixé, l'intégrande est quadratique par morceaux par rapport à f , et quadratique par rapport à ∇f . D'après les conditions imposées sur la fonction régularisante φ , les valeurs retournées par la variable b appartiennent à l'intervalle $[0; \gamma]$ et dans notre cas, la fonction φ utilisée est telle que $\gamma = 1$ (cf. tableau 2.1). La variable auxiliaire b indique la présence ou non de contours puisque $b(x) \simeq 0$ si x appartient à un contour et $b(x) \simeq 1$ sur les zones homogènes. D'après (2.20), la minimisation de J_ε par rapport à f est ramenée à celle de J_ε^* par rapport à (f, b) . Le terme $W'(f(x))$ étant linéaire par morceaux, on note :

$$W'_{local}(f(x)) = 2[A_{local}f(x) + B_{local}], \quad (2.22)$$

avec A_{local} et $B_{local} \in \mathbb{R}$. En se référant à (2.13), nous pouvons remarquer que l'expression de $W'_{local}(f(x))$, et plus précisément la valeur des paramètres A_{local} et B_{local} , dépend de la valeur de $f(x)$. En effet, selon les valeurs de $f(x)$, l'expression de $W(f(x))$ se met sous la forme P_i ou Q_i dans (2.13). En d'autres termes, l'expression de $W'_{local}(f(x))$ peut être celle de P'_i ou de Q'_i .

Par le biais des équations d'Euler-Lagrange, la minimisation de J_ε^* définie dans (2.11) par rapport à f conduit à résoudre une EDP localement linéaire, et la minimisation par rapport à b est immédiate puisque l'expression de b_{inf} est donnée par la formule (2.19).

L'algorithme de minimisation de J_ε^* lorsque $\varepsilon \rightarrow 0$ repose sur une succession de minimisations alternées comme l'indique le tableau 2.2. L'EDP ii) issue de la minimisation de J_ε^* par rapport à f (cf. tableau 2.2) est résolue par un algorithme de gradient conjugué classique [Hageman et Young, 1981, Lascaux et Theodor, 1987]. En pratique, l'initialisation f_0 de f_ε est telle que $f_0 = I$.

L'indice l est relatif à la boucle sur ε (résolution de (2.12)), alors que l'indice p est relatif aux étapes concernant les minimisations alternées de (2.21) pour chaque valeur de ε .

```

initialiser  $f_0$ 
initialiser  $\varepsilon_0$ 
• répéter (boucle sur  $l$ ) :
   $\varepsilon_l = \rho \varepsilon_{l-1}$  avec  $\rho < 1$ 
   $f_{\varepsilon_l}^0 = f_{\varepsilon_{l-1}}^{min}$ 
  ◦ répéter (boucle sur  $p$ ) :
     $f = f_{\varepsilon_l}^p$ 
     $b = b_{\varepsilon_l}^p$ 
    incrémenter  $p$ 
    i) résoudre  $\min_b J_{\varepsilon_l}^*(f, b)$  avec  $f$  fixé :
      
$$b = \frac{\varphi'(|\nabla f|)}{2|\nabla f|}$$

    ii) résoudre  $\min_f J_{\varepsilon_l}^*(f, b)$  avec  $b$  fixé :
      
$$f - \lambda^2 \varepsilon_l \operatorname{div}(b \nabla f) + \frac{\eta^2}{2\varepsilon_l} W'(f) = I$$

    ◦ jusqu'à convergence sur  $f_{\varepsilon_l}^p \rightarrow f_{\varepsilon_l}^{min}$ 
  • faire décroître  $\varepsilon_l$  jusqu'à convergence sur  $f_{\varepsilon_l}^{min}$ 

```

TAB. 2.2 – Algorithme associé à (2.12).

Les critères d'arrêt pour les deux boucles de l'algorithme indiqué dans le tableau 2.2 peuvent porter sur l'évolution relative des solutions entre deux itérations.

La discrétisation des équations i) et ii) de l'algorithme présenté dans le tableau 2.2 est donnée en annexe C.

2.6 Résultats expérimentaux

Nous présentons quelques résultats expérimentaux menés sur des données synthétiques ainsi que sur des données réelles.

Les données synthétiques intitulées "gdr" proviennent de la banque de données du GDR-PRC ISIS⁴, alors que les images SPOT ont été mises à notre disposition par le CNES.

Les résultats de classification sont donnés en fausses couleurs.

2.6.1 A propos du choix des paramètres

Il est nécessaire d'ajuster les paramètres λ et η du modèle avant l'exécution du programme dont l'algorithme est présenté dans le tableau 2.2. Le poids de chaque terme du

⁴<http://www-isis.enst.fr/>

critère (2.12) est choisi selon la nature des données, surtout selon le niveau de bruit présent sur les données et le degré de régularité souhaité sur le résultat de classification (valeur de λ). Notons que nous introduisons un paramètre δ lié à la hauteur des discontinuités dans l'image et qui divise $|\nabla f|$. Cela correspond à un changement d'échelle de $|\nabla f|$ qui s'effectue au sein du terme régularisant $\varphi(\frac{|\nabla f|}{\delta})$.

En plus des paramètres liés aux poids de termes du critère proposé, il faut fixer la valeur initiale du paramètre de convergence ε , ainsi que sa vitesse de décroissance. La valeur initiale est suffisante pour rendre le terme de contrainte de niveaux négligeable devant le terme de restauration (la valeur initiale exacte varie en fonction des données). Concernant la vitesse de décroissance de ε , nous avons opté pour une loi du type $\varepsilon_{l+1} = \rho \varepsilon_l$ (cf. tableau 2.2) avec $\rho < 1$. En pratique, la valeur de ρ est de l'ordre de 0.9 ; tout dépend du niveau de bruit des données. Pour un niveau de bruit important, il est nécessaire de restaurer l'image "plus longtemps" avant de la classer, et dans ce cas il est préférable d'opter pour une décroissance lente de ε^5 (i.e. une valeur élevée de ρ).

Au cours de la convergence sur ε , il s'avère nécessaire de diminuer le poids du terme d'attache aux données dans le cas d'images dont le niveau de bruit est important. En effet, les termes de restauration et de contrainte de niveaux tendent vers zéro, ainsi, le terme d'attache aux données risque de ramener la solution f_ε vers les données bruitées I . Il n'y a pas de relation évidente entre la loi de décroissance du terme d'attache aux données et celle du paramètre ε .

2.6.2 Données synthétiques

• Image "échiquier"

L'image synthétique présentée sur la figure 2.6 est constituée de 4 classes et nous avons introduit un bruit gaussien de moyenne nulle et additif, si bien que la valeur du RSB est de 10dB. Les paramètres gaussiens (moyennes et écart-types) des classes sont exposés dans le tableau 2.3 : l'écart-type est le même pour toutes les classes ; c'est celui du bruit gaussien ajouté aux données. On montre sur la figure 2.5 le potentiel de contrainte de niveaux W pour ces données. A partir du moment où l'écart-type est le même pour toutes les classes, tous les puits de W ont la même largeur. Nous présentons sur la figure 2.6 les résultats de classification (fonction f) et la carte des discontinuités associée (variable b) après 10 itérations sur ε . La valeur des étiquettes des classes étant celles des moyennes μ_i , nous pouvons

⁵en pratique, on observe quelques similitudes expérimentales entre le choix de la décroissance de ε et celle de la température dans un algorithme de recuit simulé.

calculer le RSB sur la solution finale, qui atteint 27.9dB. La figure 2.12 montre l'évolution du nombre de pixels mal classés, que nous appellerons "erreurs", ainsi que celle du RSB au cours de la décroissance sur ε . On constate que l'on atteint un niveau constant du RSB et du nombre d'erreurs au bout de 10 itérations sur ε . La détection de ce niveau constant peut servir de critère d'arrêt pour la boucle sur l dans l'algorithme du tableau 2.2.

Sur la figure 2.7, nous montrons l'influence du choix de la fonction régularisante sur le résultat de classification. La théorie de Van der Waals-Cahn-Hilliard concerne uniquement les fonctions quadratiques, mais les meilleurs résultats sont obtenus pour les fonctions à asymptote finie en l'infini. On constate ce phénomène sur la figure 2.7 que les fonctions non convexes et plus particulièrement avec celle proposée par Geman et McClure (cf. tableau 2.1 et [Geman et McClure, 1985]). C'est cette dernière fonction régularisante que nous utilisons pour tous les résultats présentés par la suite.

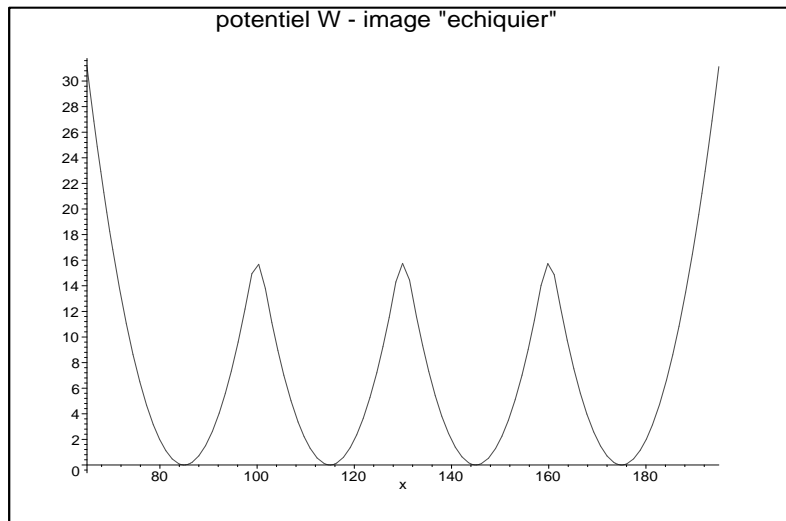


FIG. 2.5 – Potentiel W pour l'image "échiquier".

• Image "gdr"

L'image synthétique présentée sur la figure 2.10 est fournie par le GDR-PRC ISIS, et nous avons introduit un bruit gaussien additif de telle sorte que la valeur du RSB soit de 10dB. Nous avons effectué des tests en choisissant 3 classes, puis 8 classes afin d'étudier l'influence du nombre de classes sur les résultats.

Dans un premier temps, nous avons choisi 3 classes arbitraires dont les paramètres gaus-

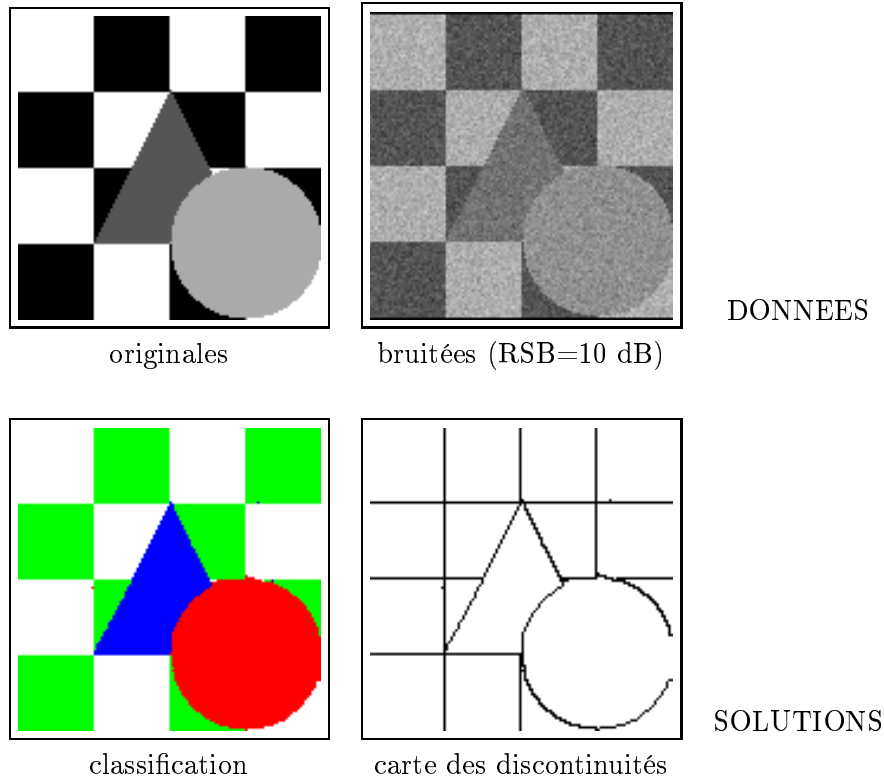


FIG. 2.6 – Résultats pour l'image "échiquier". Les paramètres utilisés sont : $\lambda = 16$, $\delta = 15$, $\eta = 0.1$, avec 7 itérations sur ε .

siens (moyennes et écart-types) sont exposés dans le tableau 2.3 : l'écart-type est le même pour toutes les classes ; c'est celui du bruit gaussien ajouté aux données. On montre sur la figure 2.8 la forme du potentiel de contrainte de niveaux W pour ces données. A partir du moment où l'écart-type est le même pour toutes les classes, les puits de W ont tous la même largeur. Le résultat de classification est présenté sur la figure 2.10, et a été obtenu après 16 itérations sur ε . Sur cette même figure nous présentons également la classification "parfaite" servant de vérité terrain pour juger de la qualité des résultats. On constate que les erreurs proviennent essentiellement de la frontière interne de la zone de dégradé où le bruit introduit rend plus difficile la détermination de cette frontière.

Nous avons ensuite choisi 8 classes arbitraires dont les paramètres sont exposés dans le tableau 2.3. On montre sur la figure 2.9 le potentiel de contrainte de niveaux W pour ces données. Remarquons que les classes de moyennes 156.0, 160.0 et 163.0 sont très proches, voire très difficiles à distinguer à travers le potentiel W . On est en présence d'un mélange de classes. Le résultat de classification est présenté sur la figure 2.11, et a été obtenu après

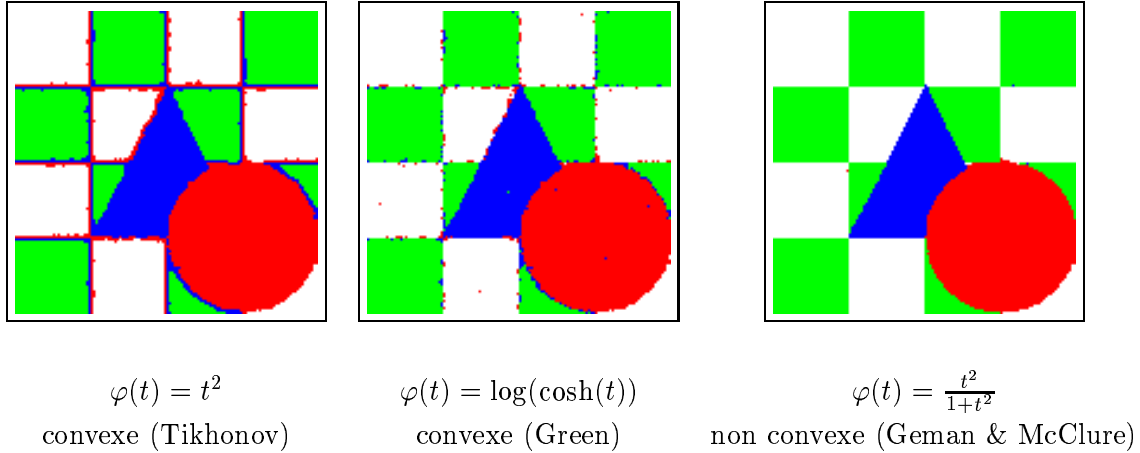


FIG. 2.7 – Classification de l'image "échiquier" à l'aide de différentes fonctions φ (voir le tableau 2.1).


























40 itérations sur ε . Sur cette même figure, nous présentons également la classification "parfaite" servant de vérité terrain pour juger de la qualité des résultats. On constate que les erreurs proviennent essentiellement des frontières internes des classes situées dans la zone dégradée et au niveau des contours de la forme irrégulière (appartenant à la classe 7) située dans le quart sud-est de l'image. Le niveau moyen d'intensité de cette forme est de 160.0, i.e. très proche de la moyenne de la classe 8 qui est l'étiquette des pixels mal classés au niveau des contours de cet objet. La figure 2.13 présente l'évolution du nombre de pixels mal classés et celle du RSB au cours de la convergence sur ε . On constate sur ces courbes qu'à partir de l'itération 40, le nombre de pixels mal classés ainsi que le RSB sont constants, d'où l'arrêt du programme à ce stade. Les figures 2.14 et 2.15 montrent l'évolution de la solution et de la carte des discontinuités associée au cours des itérations. Au cours des itérations, la valeur de ε diminue. De la première itération jusqu'à la trentième, le terme de restauration est prédominant. A partir de la trentième, ε devient suffisamment petit pour rendre le terme de contrainte de niveaux prépondérant par rapport à celui de restauration (voir (2.11)). A partir de la 33^{ième} itération, on constate clairement la restriction des niveaux de gris $f_\varepsilon(x)$ à l'ensemble des 8 étiquettes. Ce phénomène est surtout visible dans la zone de dégradé (dans le quart nord-est de l'image) où l'on voit progressivement apparaître les frontières des 6 classes présentes dans cette région de l'image. Le résultat final $f_{\varepsilon=40}$ est celui présenté en fausses couleurs sur la figure 2.11.

2.6.3 Données réelles

Nous présentons deux résultats sur des images monospectrales SPOT (bande XS3) fournies par le CNES.

Les données satellitaires de la figure 2.16 ont été prises sur une zone agricole hollandaise. Le nombre de classes et leurs paramètres gaussiens ont été estimés par un expert [Kato, 1994]. L'estimation des paramètres des 10 classes a été effectuée sur une zone située dans la partie supérieure gauche de l'image. Cette image contient 10 classes représentant des cultures de différentes natures (blé, orge,...), et les paramètres sont donnés dans la tableau 2.3. Le résultat de classification est donné en fausses couleurs sur la figure 2.16. On constate que les structures linéiques sont préservées et que les classes sont constituées de régions homogènes. Les petites régions situées en bas à droite de l'image de classification sont essentiellement dues au fait que les paramètres des classes ont été estimés sur la partie supérieure gauche des données.

Les données satellitaires SPOT de la figure 2.17 contiennent 4 classes dont les paramètres présentés dans le tableau 2.3 ont été estimés dans [Berthod *et al.*, 1996]. Le résultat de classification est donné en fausses couleurs sur la figure 2.17. L'évolution de la solution f_ε et de la carte des discontinuités associée au cours des itérations sur la décroissance de ε sont présentées sur la figure 2.18. On remarque la restriction des valeurs de f_ε à l'espace des 4 étiquettes au cours de la convergence sur ε .

CLASSES										
	1	2	3	4	5	6	7	8	9	10
image "échiquier"										
	μ_i	85.0	115.0	145.0	175.0					
	σ_i	12.85	12.85	12.85	12.85					
image "gdr" 3 classes										
	μ_i	100.0	128.0	160.0						
	σ_i	5.56	5.56	5.56						
image "gdr" 8 classes										
	μ_i	100.0	128.0	135.0	142.0	149.0	156.0	160.0	163.0	
	σ_i	5.56	5.56	5.56	5.56	5.56	5.56	5.56	5.56	
1 ^{ere} image SPOT										
	μ_i	54.6	73.5	82.5	93.8	100.5	122.8	129.9	146.6	159.9
	σ_i	9.65	2.02	5.96	9.68	17.57	2.98	6.11	3.91	5.59
2 ^{nde} image SPOT										
	μ_i	30.3	37.4	61.3	98.2					
	σ_i	2.86	2.14	11.32	11.27					
	1	2	3	4	5	6	7	8	9	10

TAB. 2.3 – Paramètres (μ_i, σ_i) et couleur de l'étiquette pour chaque classe des images traitées.

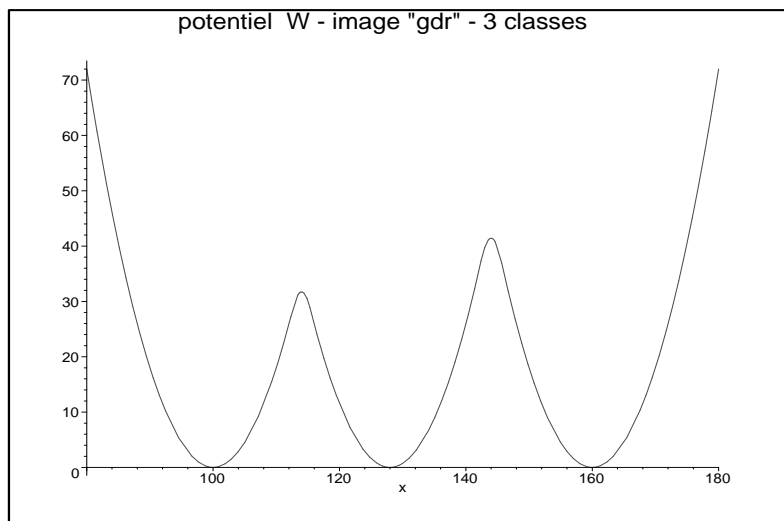


FIG. 2.8 – Potentiel W pour l'image "gdr" contenant 3 classes.

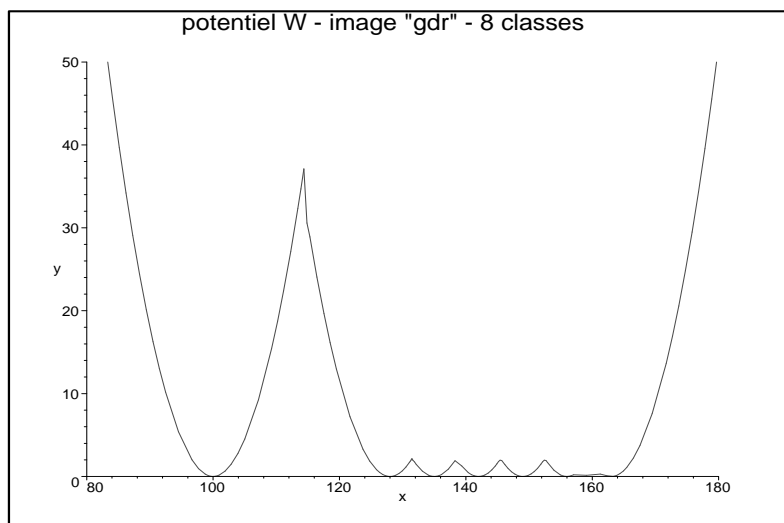


FIG. 2.9 – Potentiel W pour l'image "gdr" dans le cas de 8 classes.

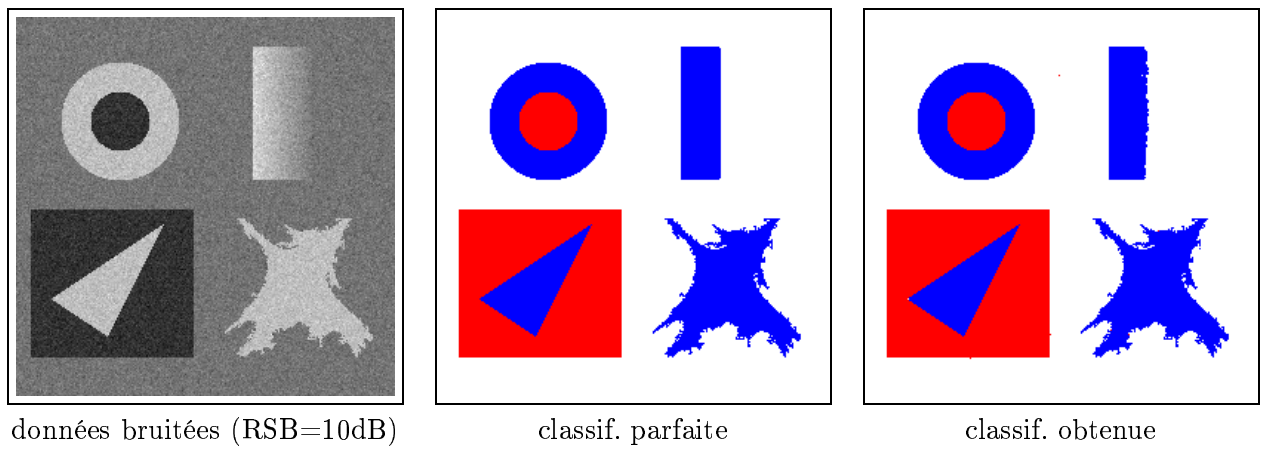


FIG. 2.10 – Résultat de classification pour l'image "gdr" et dans le cas de 3 classes. Les paramètres utilisés sont : $\lambda = 5$, $\delta = 10$, $\eta = 0.05$, avec 16 itérations sur ε .

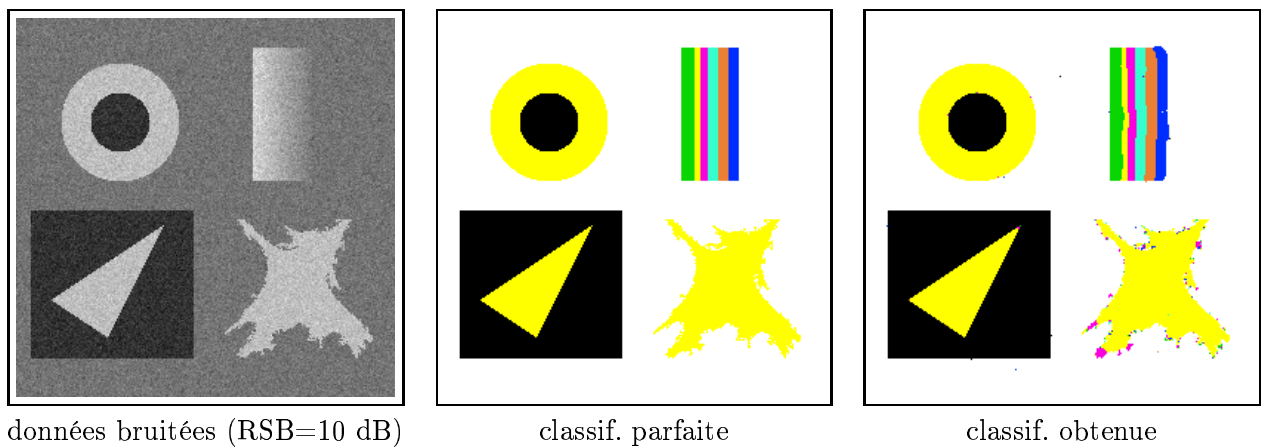


FIG. 2.11 – Résultat de classification pour l'image "gdr" dans le cas de 8 classes. Les paramètres utilisés sont : $\lambda = 60$, $\delta = 6$, $\eta = 0.05$, avec 40 itérations sur ε .

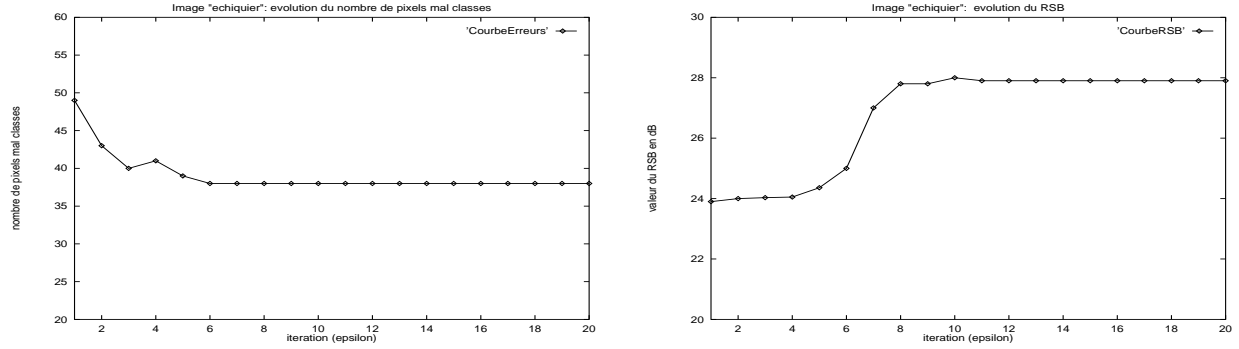


FIG. 2.12 – Image "échiquier" : évolution du nombre de pixels mal classés (à gauche) et de la valeur du RSB (à droite) en fonction des itérations sur ε .

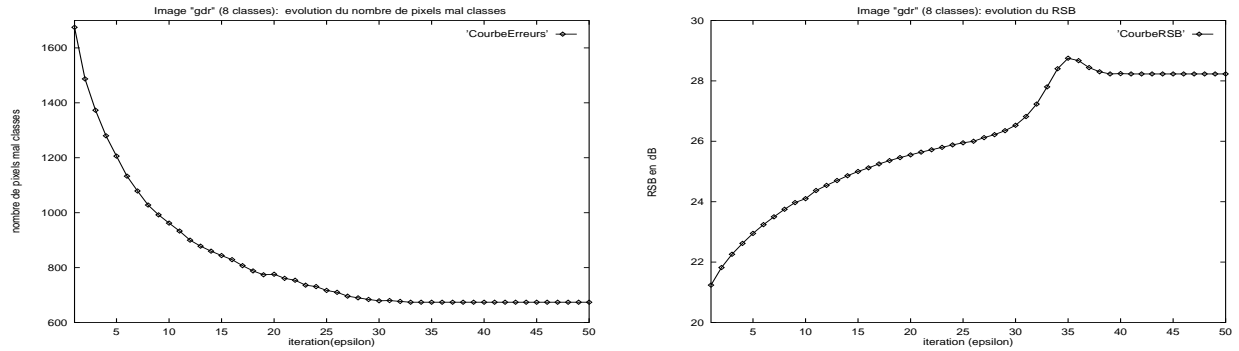


FIG. 2.13 – Image "gdr" dans le cas de 8 classes : évolution du nombre de pixels mal classés (à gauche) et de la valeur du RSB (à droite) en fonction des itérations sur ε .

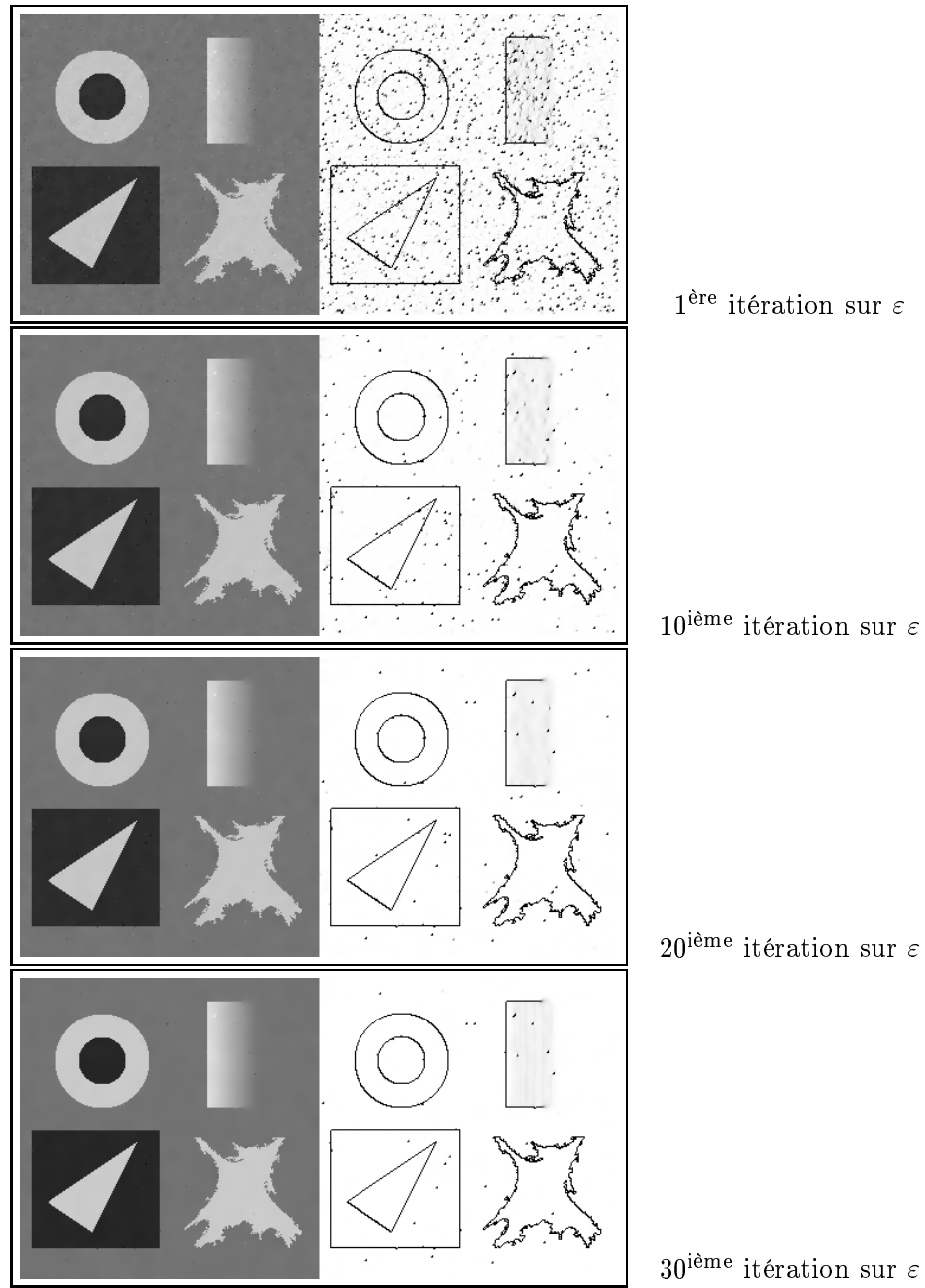


FIG. 2.14 – Image "gdr" contenant 8 classes : évolution de la solution f_ε (colonne de gauche) et de la carte des discontinuités associée (colonne de droite) lors de la décroissance de ε .

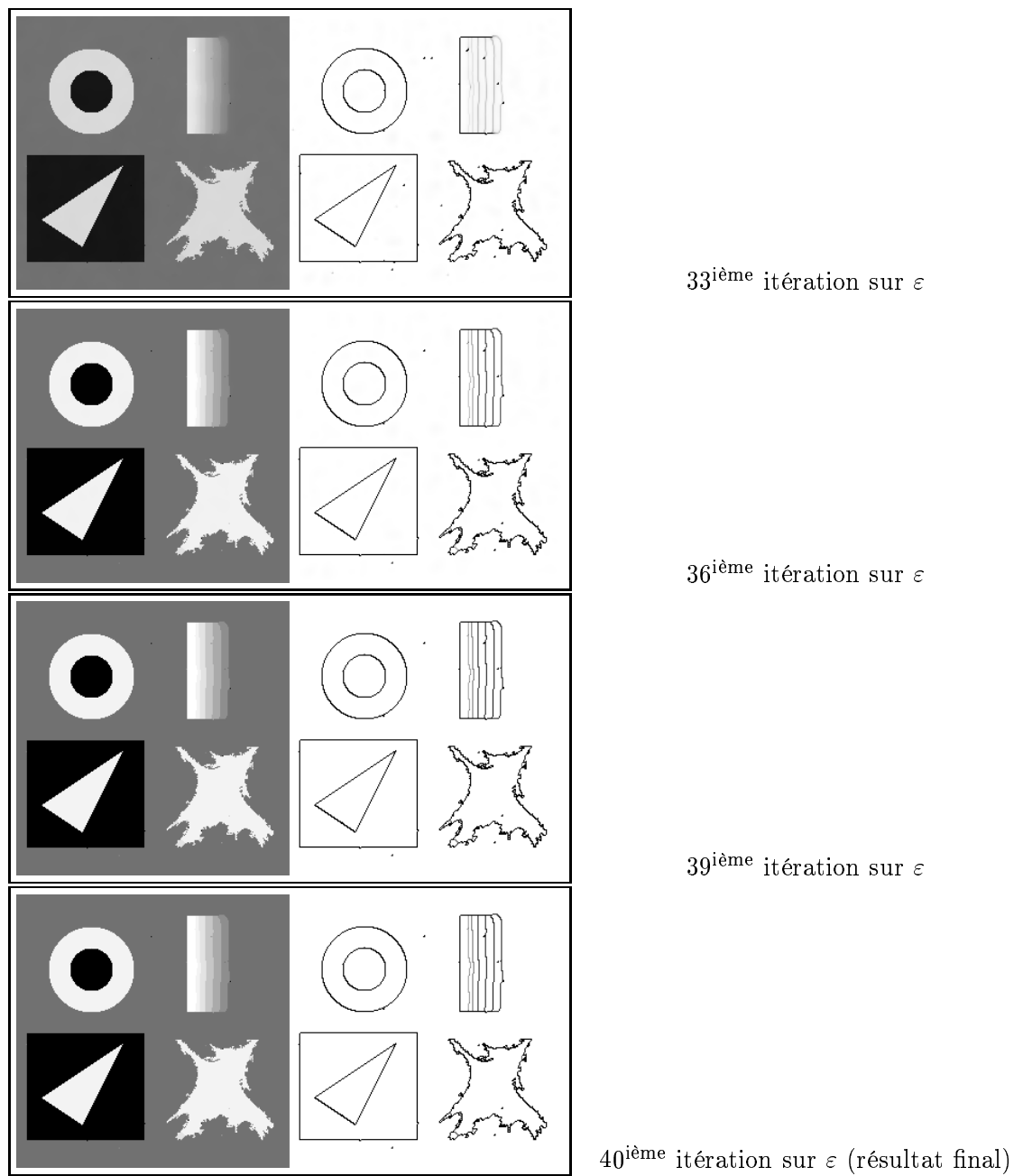


FIG. 2.15 – Suite de la figure 2.14.

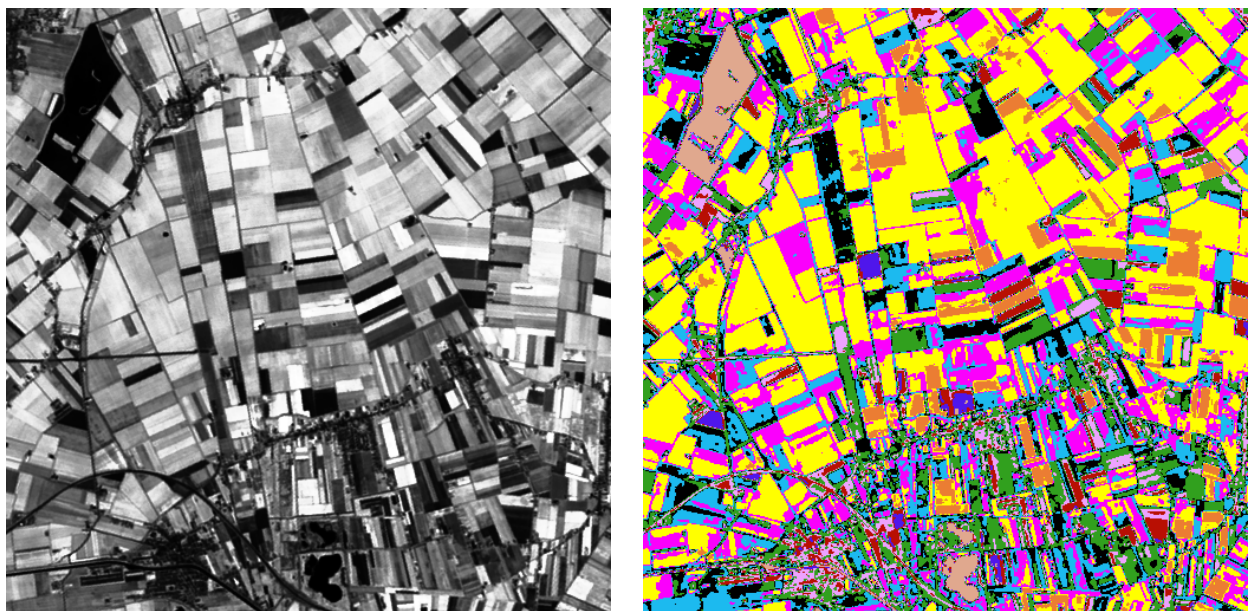


image originale

classification obtenue

FIG. 2.16 – Image originale SPOT (canal XS3) de la région hollandaise, et classification obtenue pour 10 classes.

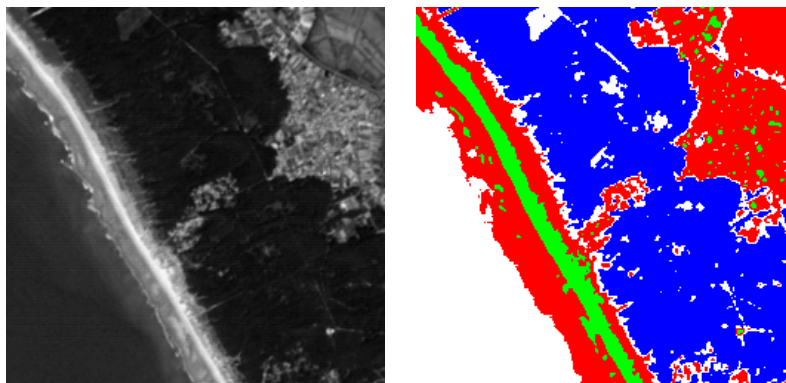


image originale

résultats obtenus

FIG. 2.17 – Image SPOT originale (canal XS3) et classification obtenue pour 4 classes.

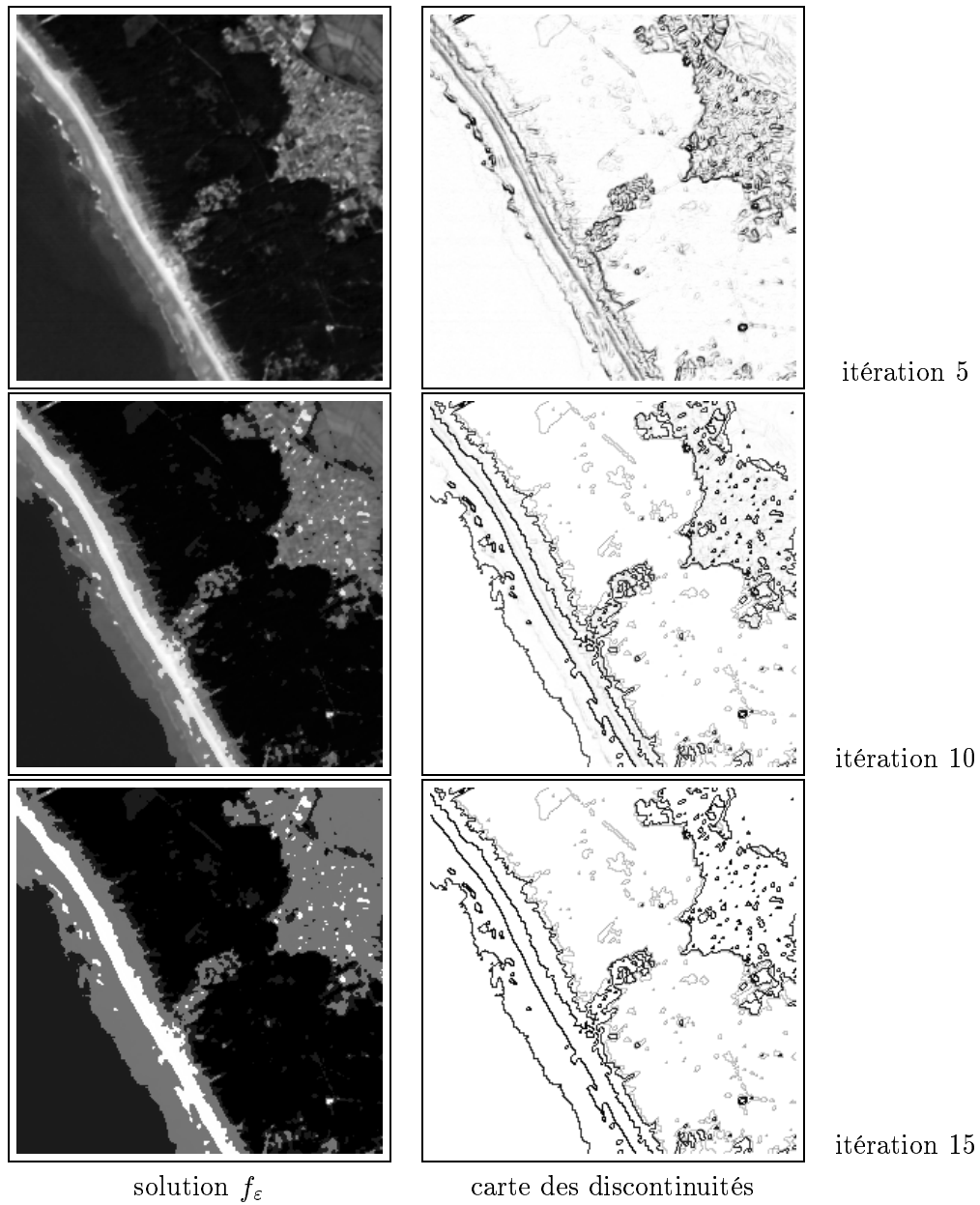


FIG. 2.18 – Seconde image SPOT contenant 4 classes. Evolution de la solution f_ϵ (colonne de gauche) et de la carte des discontinuités associée au cours des itérations sur la décroissance de ϵ .

2.7 Remarque concernant la réponse impulsionnelle

Jusqu'ici nous n'avons pas pris en compte l'effet de la réponse impulsionnelle \mathcal{R} du système d'acquisition des données. Si nous considérons l'opérateur \mathcal{R} dans le modèle de classification proposé, l'expression de la fonctionnelle de classification J_ε défini par (2.11) devient dans ce cas :

$$J_\varepsilon(f) = \int_{\Omega} \left(\mathcal{R}f(x) - I(x) \right)^2 dx + \varepsilon \lambda^2 \int_{\Omega} \varphi(|\nabla f(x)|) dx + \frac{\eta^2}{\varepsilon} \int_{\Omega} W(f(x)), \quad (2.23)$$

et le problème associé consiste à trouver \tilde{f} vérifiant :

$$\tilde{f} = \lim_{\varepsilon \rightarrow 0^+} \left[\arg \min_f J_\varepsilon(f) \right]. \quad (2.24)$$

Nous utilisons également la méthode de *régularisation semi-quadratique* [Charbonnier *et al.*, 1997, Geman et Reynolds, 1992] en introduisant une variable auxiliaire b liée à la présence de contours dans l'image (cf. paragraphe 2.5). Nous obtenons ainsi la fonctionnelle (appelée *énergie augmentée*) suivante :

$$J_\varepsilon^*(f, b) = \int_{\Omega} \left(\mathcal{R}f(x) - I(x) \right)^2 dx + \varepsilon \lambda^2 \int_{\Omega} \left[b(x) |\nabla f(x)|^2 + \psi(b(x)) \right] dx + \frac{\eta^2}{\varepsilon} \int_{\Omega} W(f(x)) dx. \quad (2.25)$$

Pour chaque valeur de ε , nous opérons alors en alternance une minimisation de J_ε^* par rapport à b avec f fixé, puis par rapport à f pour b fixé. Lorsque ε est fixé, l'équation d'Euler-Lagrange associée à la minimisation de J_ε^* par rapport à f conduit à l'EDP :

$$\mathcal{R}^* \mathcal{R} f - \lambda^2 \varepsilon \operatorname{div}(b \nabla f) + \frac{\eta^2}{2\varepsilon} W'(f) = \mathcal{R}^* I. \quad (2.26)$$

L'algorithme de minimisation utilisé est identique à celui présenté dans le tableau 2.2 du paragraphe 2.5, en remplaçant l'EDP relative à la minimisation de J_ε^* par rapport à f par (2.26).

Nous montrons sur la figure 2.19 un exemple de classification obtenue sur des données synthétiques floues (filtre gaussien avec un noyau de convolution de taille 11×11 pixels, de moyenne nulle et de variance 1.19).

2.8 Conclusion

Le modèle proposé dans ce chapitre permet de réaliser simultanément une restauration ainsi qu'une classification de l'image. Nous avons défini une famille de fonctionnelles dont

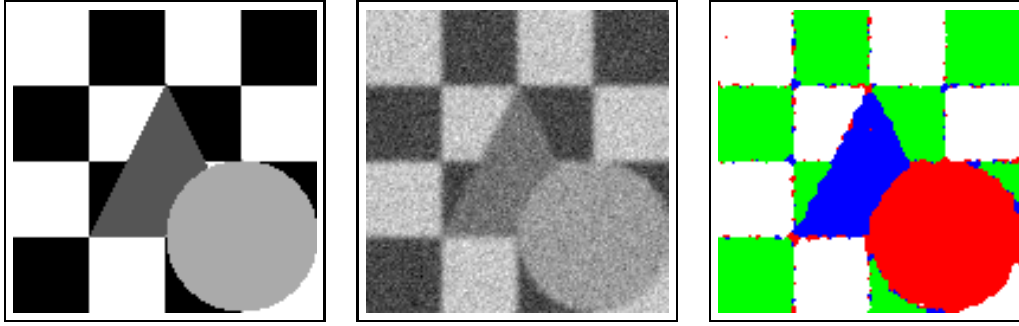


FIG. 2.19 – Image synthétique floue et bruitée (RSB=10dB) et classification obtenue d’après le modèle variationnel donné par (2.23).

la suite de solutions converge progressivement vers une image restaurée avec préservation des contours (terme de régularisation), puis vers un étiquetage des pixels par attraction des niveaux de l’intensité vers l’ensemble des étiquettes (potentiel de contraintes de niveaux). Les fondements théoriques de cette approche reposent sur les travaux conduits en mécanique (surtout en mécanique des fluides) ainsi qu’en mathématiques (essentiellement fondés sur la Γ -convergence des suites de fonctionnelles) dans le cadre de la théorie de Van der Waals-Cahn-Hilliard. Nous avons adapté ces travaux au cadre de la classification d’image. Il en résulte un problème ouvert concernant l’extension des résultats théoriques lors de l’utilisation de perturbations sous-quadratiques qui fournissent les meilleurs résultats en ce qui concerne la préservation des contours (aspect restauration) et le nombre de pixels mal classés (aspect classification). L’algorithme résultant est efficace comme l’illustrent les expérimentations menées aussi bien sur des données synthétiques que réelles. Sa mise en œuvre est effectuée à l’aide du principe de régularisation semi-quadratique permettant de s’affranchir de la non convexité du critère à minimiser.

Par souci de simplicité, nous avons tenu à développer la théorie dans le cas où la réponse impulsionnelle du système n’est pas prise en compte, puis nous l’avons introduite à la fin du chapitre en l’illustrant d’un test expérimental.

Nous présentons la version multispectrale de ce modèle au sein du chapitre 5. Les points délicats lors de l’extension au cas de données multiples, proviennent de la prise en compte de l’information des différentes bandes spectrales à travers le calcul du module du gradient au sein de la fonction régularisante, ainsi qu’à travers le potentiel de contrainte de niveaux qui devra être défini sur un espace multi-dimensionnel.

Chapitre 3

Classification par modèle dynamique



Où nous proposons un second modèle de classification mettant en jeu un ensemble de contours et de régions actives dont l'évolution est guidée par des forces internes et externes. L'approche est différente de la précédente dans le sens où le problème n'est plus posé sous la forme d'un problème de restauration sous contrainte de classification, mais sous la forme d'un problème de partitionnement des données observées en fonction des classes prédéfinies. L'originalité de cette approche repose sur le fait que la partition optimale cherchée est définie à l'aide d'ensembles de niveaux. Après un rappel succinct sur les travaux antérieurs effectués dans ce domaine, nous établissons les propriétés de régularité souhaitées sur la partition finale, puis nous utilisons une formulation par ensembles de niveaux pour modéliser les classes. Nous obtenons un critère que nous minimisons à l'aide d'un schéma dynamique. Nous terminons ce chapitre par des applications sur des images synthétiques et réelles.

3.1 Travaux antérieurs

Nous passons en revue quelques travaux et approches concernant la segmentation d'images par contours actifs. Les articles [Caselles *et al.*, 1993, Caselles *et al.*, 1997, Kichenassamy *et al.*, 1996] contiennent de plus amples informations et références quant aux différents travaux relatifs à ce sujet. Nous abordons également la formulation par ensembles de niveaux permettant d'implémenter les contours actifs tout en gérant les changements de topologie des courbes. Enfin, nous présentons un certain nombre de travaux liés à la segmentation par modèles actifs fondés sur des critères de type région.

3.1.1 Les contours actifs et formulation par ensembles de niveaux

Les précurseurs concernant la détection d'objets à partir de contours actifs furent Kass *et al.* [Kass *et al.*, 1987]. Leurs travaux ont conduit à l'élaboration des *modèles classiques* qui sont à l'origine de l'effort de recherche porté dans le domaine des modèles déformables. Dans le cadre des modèles classiques, un contour actif, *snake* en Anglais, résulte de la minimisation d'une énergie, de telle sorte qu'à travers les différents termes du critère énergétique, le minimum est atteint lorsque le contour déformable se trouve sur les parois de l'objet à détecter et vérifie des contraintes de régularité géométrique.

Dans le cas 2D, notons $C : [a, b] \longrightarrow \mathbb{R}^2$ la courbe paramétrée à laquelle est associée l'énergie E suivante :

$$E(C(q)) = \alpha \int_a^b |C'(q)|^2 dq + \beta \int_a^b |C''(q)|^2 dq - \lambda \int_a^b |\nabla I(C(q))|^2 dq \quad (3.1)$$

où α , β , et λ sont des constantes réelles positives, et I représente l'intensité de l'image à segmenter. Les méthodes classiques offrent un intérêt certain dans le cadre de la segmentation d'images, surtout par l'innovation que représente l'introduction des contours déformables. Cependant, cette approche comporte les inconvénients suivants :

- L'impossibilité de gérer un changement de topologie : l'approche classique ne permet pas de segmenter plusieurs objets à partir d'un seul contour.
- La formulation du problème dépend de la paramétrisation de la courbe choisie.
- Le contour initial doit être suffisamment proche de l'objet pour pouvoir converger vers celui-ci, au risque de tomber dans un minimum local.
- La gestion des points de la courbe induit des problèmes d'échantillonnage.

Ces aspects négatifs ont conduit à améliorer les modèles classiques pour s'affranchir des inconvénients qui leur sont rattachés.

Malladi *et al.* [Caselles *et al.*, 1993, Malladi *et al.*, 1994, Malladi *et al.*, 1995], ont largement contribué à l'étude des modèles déformables grâce à leur travaux sur les *contours actifs géométriques*. Leur approche se distingue de celle suivie par Kass *et al.* par le fait que l'évolution du contour actif est perçue comme une propagation de front d'onde. Ces auteurs définissent une EDP dynamique gouvernant l'évolution du contour déformable $C(q; t)$ où t représente le temps. Cette EDP ne résulte pas de la minimisation d'un critère énergétique comme précédemment. L'expression générale de l'EDP dynamique introduite ici est la suivante :

$$\begin{cases} \frac{\partial C(q; t)}{\partial t} = F \vec{N} \\ C(q, 0) = C_0(q) \end{cases} \quad (3.2)$$

où \vec{N} est la normale unitaire intérieure à C au point q et au temps t , et F est une vitesse dont les termes sont liés aux propriétés géométriques du contour. A $t = 0$, le contour actif est initialisé par la courbe $C_0(q)$. Si F est la courbure moyenne κ du contour, alors l'EDP de (3.2) conduit à une évolution du contour par courbure moyenne (*mean curvature motion* en Anglais). Ce type d'EDP possède un certain nombre de propriétés intéressantes, comme la régularisation de la courbe par minimisation de longueur, et a suscité de nombreux travaux en traitement d'images (voir [Kichenassamy *et al.*, 1996] pour plus de détails).

Par ailleurs, Caselles *et al.* ainsi que Malladi *et al.* ont introduit les contours actifs **géodésiques** [Caselles *et al.*, 1993, Caselles *et al.*, 1997, Malladi *et al.*, 1994]. Ils ont défini l'EDP dynamique suivante :

$$\frac{\partial C}{\partial t} = g(|\nabla I|)[\kappa + c]\vec{N} - (\nabla g \cdot \vec{N})\vec{N}. \quad (3.3)$$

où g est une fonction permettant d'arrêter l'évolution du contour sur les zones dont le module du gradient de l'intensité de l'image observée I est élevé (ces zones sont supposées être liées à la présence de contours d'objets à segmenter). La fonction $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ est monotone décroissante et vérifie $g(0) = 1$, $\lim_{x \rightarrow \infty} g(x) = 0$. Il est courant d'utiliser une version lisse des données I pour calculer $g(|\nabla I|)$ et éviter ainsi les forts gradients dus au

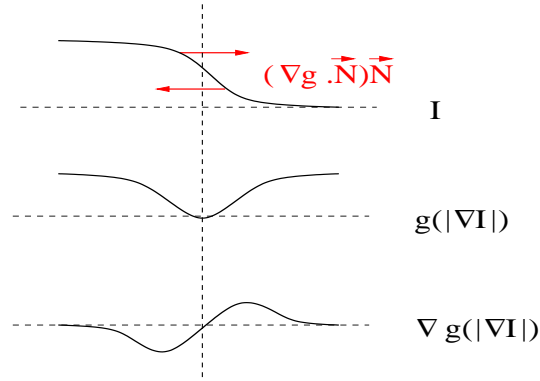


FIG. 3.1 – Influence de la fonction d'arrêt $g(|\nabla I|)$ pour $g(t) = \frac{1}{1+t^2}$.

bruit. L'influence de la fonction d'arrêt g et du terme $(\nabla g \vec{N}) \vec{N}$ est illustrée sur la figure 3.1. La constante réelle c est appelée *force ballon* [Cohen, 1991] induisant une force constante sur les parois du contour (inflation ou contraction du contour) qui permet d'imprimer un mouvement régulier et également de compenser un changement de signe du terme de courbure κ lors d'un passage de la courbe d'un état convexe à un état non convexe par exemple.

Afin de palier les problèmes relatifs à la gestion du changement de topologie des contours, on utilise une formulation du problème par **ensembles de niveaux** proposée par [Osher et Sethian, 1988]. L'ouvrage [Sethian, 1996] présente en détails cette approche. L'idée consiste à représenter le contour $C(q; t)$ par l'ensemble de niveau (en général le niveau zéro) d'une fonction $u : \mathbb{R}^2 \rightarrow \mathbb{R}$ lipschitzienne continue et dépendant du temps t (voir la figure 3.2). Ainsi, l'évolution du contour C est représentée par l'évolution du niveau zéro de la fonction u et noté $\Gamma(t) = \{x \in \Omega / u(x; t) = 0\}$. De plus, afin de distinguer la région intérieure $\Omega_{int}(t)$ de la région extérieure $\Omega_{ext}(t)$ délimitée par la courbe fermée $\Gamma(t)$, la fonction u est généralement signée de la façon suivante :

$$\begin{cases} u(x; t) = 0 & \text{si } x \in \Gamma(t) \\ u(x; t) = \epsilon d(x, \Gamma(t)) & \text{si } x \in \Omega_{int}(t) \\ u(x; t) = -\epsilon d(x, \Gamma(t)) & \text{si } x \in \Omega_{ext}(t) \end{cases} \quad (3.4)$$

avec $\epsilon = \pm 1$ et $d(x, \Gamma(t)) = \inf_{y \in \Gamma(t)} \|x - y\|_{\mathbb{R}^2}$.

Lorsque $\epsilon = -1$, l'équation d'évolution (3.2) devient dans le cadre d'une formulation par

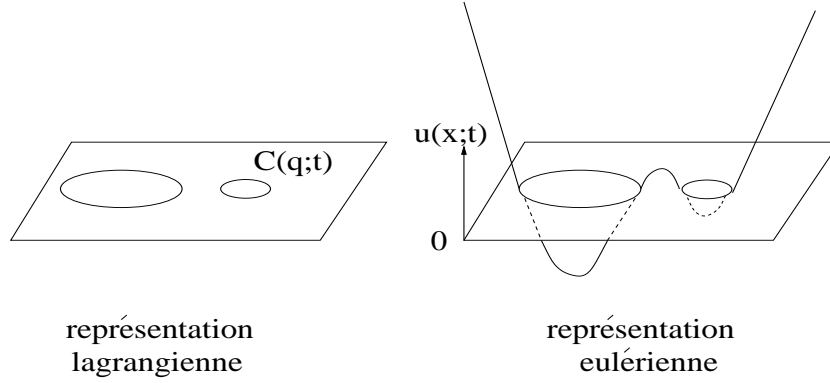


FIG. 3.2 – Représentation du contour C par l'ensemble de niveau zéro d'une fonction u .

ensembles de niveaux¹ :

$$\frac{\partial u}{\partial t} = F|\nabla u| \quad (3.5)$$

pour $\vec{N} = -\frac{\nabla u}{|\nabla u|}$. Cette nouvelle formulation permet un changement de topologie de la courbe active comme la division ou la fusion des contours. La formulation par ensemble de niveaux rend ainsi possible la détection de deux objets à partir d'un unique contour actif initial. L'évolution du contour d'après l'équation (3.5) dépend du terme $|\nabla u|$. On initialise souvent la fonction u telle que

$$\begin{cases} u(x;t) = 0 & \text{si } x \in \Gamma(t) \\ u(x;t) < 0 & \text{si } x \in \Omega_{int}(t) \\ u(x;t) > 0 & \text{si } x \in \Omega_{ext}(t) \end{cases} \quad (3.6)$$

et vérifiant $|\nabla u| = 1$. Notons alors que toutes les courbes de niveaux de u sont solutions de (3.5), ce qui engendre une perte de régularité de la fonction distance u : le module du gradient de u tend vers l'infini au cours du temps. Lorsque $F = \kappa$, la courbure du contour, il est montré dans [Barles *et al.*, 1993, Evans et Spruck, 1992] que la fonction de distance n'est pas solution de (3.5). Afin de préserver numériquement la contrainte $|\nabla u| = 1$, les auteurs de [Sussman *et al.*, 1994] ont proposé d'appliquer régulièrement sur la fonction u l'EDP suivante :

$$\frac{\partial u}{\partial t} = \text{signe}(u)(1 - |\nabla u|) \quad (3.7)$$

¹ $C(q;t)$ est représenté par $\{x = (x_1, x_2) \in \Omega / u(x;t) = 0\}$

$du = \frac{\partial u}{\partial t} + \frac{\partial u}{\partial x_1} \frac{\partial x_1}{\partial t} + \frac{\partial u}{\partial x_2} \frac{\partial x_2}{\partial t} = 0$

$\frac{\partial u}{\partial t} + \nabla u \cdot \kappa \vec{N} = \frac{\partial u}{\partial t} - \kappa |\nabla u| = 0$ car $\vec{N} = -\frac{\nabla u}{|\nabla u|}$

avec

$$\text{signe}(u) = \begin{cases} +1 & \text{si } u > 0 \\ -1 & \text{si } u < 0 \\ 0 & \text{si } u = 0 \end{cases}$$

Cette EDP permet de maintenir la contrainte $|\nabla u| = 1$ sur la fonction de distance signée u . Notons qu'il existe des travaux visant à maintenir la fonction u comme une fonction de distance par l'introduction de la contrainte $|\nabla u| = 1$ dans le schéma d'évolution (3.5) comme cela est proposé dans [Gomes et Faugeras, 1999, Gomes et Faugeras, 2000] et étudié ensuite dans [Aujol, 1999].

L'expression de l'équation (3.3) à l'aide d'une formulation par ensembles de niveaux devient :

$$\frac{\partial u}{\partial t} = g(|\nabla I|)[\kappa + c]|\nabla u| - \nabla g \nabla u. \quad (3.8)$$

avec $\kappa = \text{div} \left(\frac{\nabla u}{|\nabla u|} \right)$.

3.1.2 Les modèles actifs reposant sur une approche de type région

Nous allons passer en revue un certain nombre de travaux effectués dans le cadre général de la segmentation et mettant en œuvre des modèles actifs fondés sur des approches de type région. Parmi ces travaux, nous trouvons également des modèles actifs combinant l'information au niveau des contours aux statistiques des régions délimitées par ces contours.

Zhu et Yuille [Zhu et Yuille, 1996] ont proposé un modèle non supervisé de classification impliquant un ensemble de contours actifs noté Γ dont l'évolution est guidée par des contraintes géométriques (minimisation de la longueur des contours), et par un terme portant sur les propriétés statistiques des régions associées aux contours. La solution obtenue est une partition en régions $\{\Omega_i\}_{i=1\dots K}$ de l'image de support Ω telle que : $\Omega = \bigcup_{i=1}^K \Omega_i$ avec $\Omega_i \cap_{i \neq j} \Omega_j = \emptyset$. Le critère qu'ils proposent de minimiser alternativement par rapport à Γ l'ensemble des contours, et par rapport à $\{\alpha_i\}_{i=1\dots K}$ l'ensemble des paramètres inconnus caractérisant la distribution de l'intensité dans les classes, est le suivant :

$$F(\Gamma, \{\alpha_i\}) = \sum_{i=1}^K \left\{ \frac{\mu}{2} \int_{\partial\Omega_i} ds - \log P(\{I(x) : x \in \Omega_i\} / \alpha_i) + \lambda \right\},$$

où $I(x)$ est l'intensité du point x , $\partial\Omega_i$ est la frontière de la région Ω_i formée des pixels de la $i^{\text{ème}}$ classe. $P(\{I(x) : x \in \Omega_i\} / \alpha_i)$ représente la probabilité d'avoir une valeur d'intensité I en x , sachant que sur la région Ω_i la loi de distribution de I est caractérisée par les paramètres α_i . Le paramètre λ est la longueur de code nécessaire pour décrire la distribution

de la région Ω_i , λ est supposé être identique pour toutes les régions. Ainsi, par le biais de minimisations alternées, leur modèle permet d'estimer, à chaque itération, les paramètres des classes d'une part et de faire évoluer l'ensemble des contours d'autre part. Cependant, les contours étant modélisés point par point (représentation lagrangienne), il leur est plus complexe de gérer un changement de topologie des courbes actives, comme le permet aisément une formulation par ensembles de niveaux².

Yezzi *et al.* [Yezzi *et al.*, 1999] se sont inspirés de ces travaux et ont reformulé le problème à l'aide d'ensembles de niveaux mais uniquement pour le cas de deux et trois classes.

Chan et Vese [Chan et Vese, 1999] ont introduit un modèle non supervisé fondé sur une approche région et formulé dans le cadre des ensembles de niveaux. Cependant, ce modèle est restreint au cas de deux classes, définies par les régions Ω_1 et Ω_2 de niveaux moyens c_1 et c_2 , et séparées par un ensemble de frontières $\Gamma = \partial\Omega_1 \cap \partial\Omega_2$. Soit u la fonction définie par (3.4) avec $\varepsilon = +1$ telle que $\Omega_1 = \{x/u(x) > 0\}$, $\Omega_2 = \{x/u(x) < 0\}$ et $\Gamma = \{x/u(x) = 0\}$. L'expression de la fonctionnelle à minimiser alternativement par rapport aux trois variables, u , c_1 et c_2 est :

$$F(u, c_1, c_2) = \underbrace{\mu \int_{\Omega} \delta(u) |\nabla u|}_{\text{longueur de } \Gamma} + \underbrace{\nu \int_{\Omega} H(u)}_{\text{aire de } \Omega_1} + \underbrace{\lambda_1 \int_{\Omega} |I - c_1|^2 H(u)}_{\text{sur } \Omega_1 : I(x) \simeq c_1} + \underbrace{\lambda_2 \int_{\Omega} |I - c_2|^2 (1 - H(u))}_{\text{sur } \Omega_2 : I(x) \simeq c_2}$$

Les fonctions δ et H sont respectivement les distributions de Dirac et de Heaviside. Les paramètres μ, ν, λ_1 et λ_2 sont des réels fixés permettant de contrôler le poids de chaque terme. Outre l'évolution de l'ensemble des contours Γ , leur modèle permet d'estimer les paramètres c_1 et c_2 caractérisant les deux classes. La forme générale de cette fonctionnelle comporte des termes similaires à celle que nous allons proposer, comme le terme de longueur des contours par exemple, que nous expliciterons plus en détails par la suite.

Paragios et Deriche [Paragios et Deriche, 1999, Paragios, 2000] ont développé un modèle de contours actifs combinant l'information de type contour à celle de type régions pour la segmentation d'images. La partition $\mathcal{P}(\Omega)$ de Ω est obtenue en minimisant le critère :

$$F(\mathcal{P}(\Omega)) = \alpha \sum_{i=1}^K \int_{\Omega_i} -\log(p_i(I(x))) dx + (1 - \alpha) \sum_{i=1}^K \int_0^1 g(p_{B_i}(\partial\Omega_i(c_i)), \sigma_B) |\partial\dot{\Omega}_i(c_i)| dc_i.$$

$p_i(I(x))$ est la probabilité a priori de la $i^{\text{ème}}$ classe (distribution gaussienne), $\partial\Omega_i(c_i)$ est la frontière de la région Ω_i de paramètre c_i , $p_{B_i}(\partial\Omega_i(c_i))$ est la probabilité de se trouver

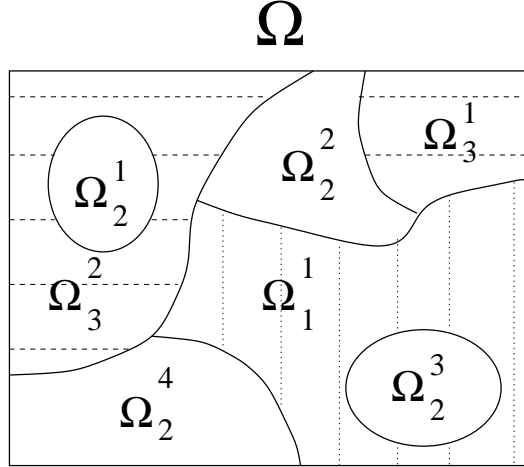
²La gestion d'un changement de topologie est relativement complexe au niveau algorithmique dans le cadre d'une formulation Lagrangienne, et "transparente" dans une formulation Eulérienne.

sur la frontière de la région Ω_i (liée par exemple aux zones où le module du gradient de l'intensité est important), et $g(., \sigma_B)$ est une gaussienne de moyenne nulle et d'écart-type σ_B . Le terme $|\partial \dot{\Omega}_i(c_i)|$ contraint la courbe à garder une certaine régularité morphologique et $\partial \dot{\Omega}_i(c_i) = \frac{\partial}{\partial c_i}[\partial \Omega_i](c_i)$. Le nombre de classes K n'est pas limité contrairement à [Chan et Vese, 1999] ou [Yezzi *et al.*, 1999]. La méthode des ensembles de niveaux est introduite au niveau des EDP résultant des équations d'Euler-Lagrange utilisées pour la minimisation du critère qu'ils proposent. Contrairement aux travaux précédemment cités, leur modèle est supervisé et la détermination des caractéristiques des classes est effectuée au préalable par un algorithme d'estimation des paramètres de texture et du nombre de classes. Un terme de couplage est introduit au niveaux des EDP pour permettre de guider l'évolution de chaque ensemble de régions associé à une classe en fonction des autres classes.

Le modèle que nous proposons [Samson *et al.*, 2000a, Samson *et al.*, 1999a, Samson *et al.*, 1999c] s'inspire des travaux effectués par Zhao *et al.* [Zhao *et al.*, 1996] sur l'évolution de phases multiples dans le cadre de la mécanique des fluides impliquant un ensemble de régions de caractéristiques différentes (les *phases*) modélisées à l'aide d'ensembles de niveaux. A chaque phase, vue comme une classe, est attribuée une fonction d'ensemble de niveaux permettant de modéliser la totalité des régions et des contours formant cette classe. Ainsi, la formulation par ensembles de niveaux est introduite directement au niveau de la fonctionnelle et non au niveau des équations d'Euler-Lagrange pour permettre un changement de topologie des contours actifs (comme c'est le cas dans [Chan et Vese, 1999, Paragios, 2000, Paragios et Deriche, 1999, Yezzi *et al.*, 1999]). La fonctionnelle proposée est définie sur ces fonctions de niveaux. Cette formulation permet de définir aisément un terme couplant l'évolution des ensembles de niveaux les uns par rapport aux autres et pénalisant les pixels sans étiquette ainsi que le chevauchement de régions. La fonctionnelle contient trois termes, le premier prend en compte les caractéristiques des classes, le second, pondéré par une fonction d'arrêt liée à la présence de frontières d'un objet de la scène, porte sur la régularité géométrique des contours, et le troisième est un terme de couplage pénalisant les régions vides (sans étiquette) et le chevauchement de régions.

3.2 La classification en tant que problème de partition

Nous allons exposer ici les propriétés que nous désirons obtenir sur la partition cherchée à l'aide du modèle de classification proposé. Soit I la fonctionnelle liée aux données observées et définie sur le support Ω que nous cherchons à partitionner. Cette partition doit être constituée de régions homogènes auxquelles ont été assignées des étiquettes, et

FIG. 3.3 – Une partition de Ω .

séparées par des interfaces régulières. On appelle **classe** un ensemble de régions possédant la même étiquette. Comme précédemment nous supposons que l'intensité I est répartie de façon gaussienne sur chaque classe et que μ_i, σ_i sont la moyenne et l'écart-type caractérisant la classe C_i .

Remarque : tous les indices i et j vont de 1 à K , K étant le nombre de classes, lorsque cela n'est pas précisé.

On note Ω_i l'ensemble des régions formant la classe C_i tel que :

$$\Omega_i = \{x \in \Omega / x \text{ appartient à la classe } C_i\}, \quad (3.9)$$

et nous avons :

$$\Omega_i = \bigcup_{j=1}^{NR_i} \Omega_i^j \quad \text{avec} \quad \Omega_i^n \cap_{n \neq m} \Omega_i^m = \emptyset, \quad (3.10)$$

où NR_i est le nombre de régions formant la classe C_i . Par exemple la classe 2 de la figure 3.3 est constituée de quatre régions ($NR_2 = 4$).

Partitionner Ω consiste à déterminer l'ensemble $\{\Omega_i\}_{i=1 \dots K}$ (voir la figure 3.3) tel que :

$$\Omega = \bigcup_{i=1}^K (\Omega_i \cup \Gamma_i) \quad \text{et} \quad \Omega_i \cap_{i \neq j} \Omega_j = \emptyset. \quad (3.11)$$

On appelle $\Gamma_i = \partial\Omega_i \cap \Omega$ l'intersection de l'ensemble des frontières de Ω_i avec l'ouvert Ω .

L'interface entre Ω_i et Ω_j est notée par :

$$\Gamma_{ij} = \Gamma_{ji} = \Gamma_i \cap \Gamma_j, \quad \forall i \neq j. \quad (3.12)$$

De plus, nous avons :

$$\Gamma_i = \bigcup_{j \neq i} \Gamma_{ij}. \quad (3.13)$$

Remarquons que dans (3.12) il est possible d'avoir $\Gamma_{ij} = \emptyset$.

On note $|\Gamma_i| = \mathcal{H}^1(\Gamma_i)$ la mesure uni-dimensionnelle de Hausdorff de l'ensemble Γ_i vérifiant :

$$|\Gamma_i| = \sum_{j \neq i} |\Gamma_{ij}| \text{ avec } |\emptyset| = 0. \quad (3.14)$$

A partir de ces définitions, nous allons énumérer les conditions que l'on souhaite imposer sur la solution du modèle de classification proposé. Le modèle considéré doit permettre de partitionner les données collectées, et représentées par la fonction I définie sur l'ouvert borné Ω , de telle sorte que la partition obtenue soit constituée de l'ensemble des régions $\{\Omega_i\}_{i=1\dots K}$ définies par (3.9) et satisfaisant les conditions exposées dans le tableau 3.1.

Les conditions A, B et C du tableau 3.1 constituent les propriétés de la solution du modèle de classification que nous allons proposer par la suite. La définition du modèle est effectuée en associant une fonctionnelle à l'ensemble des interfaces telle que la minimisation de cette fonctionnelle conduise à une solution respectant ces trois conditions.

Les paramètres ξ_{ij} de la condition C du tableau 3.1 sont fixés et permettent de prendre en compte une information *a priori* sur la longueur relative des contours les uns par rapport aux autres. Par exemple, si nous savons par avance que les contours associés à une classe possèdent une forte courbure il est utile de pouvoir diminuer la contrainte de longueur qui leur est imposée.

3.3 Formulation par ensembles de niveaux

Le modèle de classification développé plus loin implique un ensemble de régions et contours actifs couplés et l'approche adoptée s'inspire des travaux effectués par Zhao et *al.* [Zhao et al., 1996]. L'évolution de chaque contour est gouvernée par des forces contraignant la solution de classification à respecter les conditions A, B et C présentées dans le tableau 3.1. Les régions Ω_i et les contours Γ_i associés à chaque classe C_i sont modélisés à l'aide d'une fonction d'ensemble de niveaux.

CONDITION A

$\{\Omega_i\}_{i=1\dots K}$ est une partition de Ω :

$$\Omega = \bigcup_i (\Omega_i \cup \Gamma_i) \quad \text{avec} \quad \Omega_i \cap_{i \neq j} \Omega_j = \emptyset.$$

CONDITION B

La partition $\{\Omega_i\}_{i=1\dots K}$ doit prendre en compte les caractéristiques de chaque classe, à savoir la moyenne et la variance des classes compte tenu des hypothèses formulées sur la distribution (gaussienne) de l'intensité pour chaque classe. Cette condition est traduite à travers la minimisation du terme suivant :

$$\text{minimiser} \sum_i \int_{\Omega_i} \left(\frac{I(x) - \mu_i}{\sigma_i} \right)^2 dx \quad \text{par rapport aux variables } \Omega_i.$$

CONDITION C

La partition obtenue doit être régulière dans le sens où la longueur totale des interfaces Γ_{ij} est minimale :

$$\text{minimiser} \sum_{i,j} \xi_{ij} |\Gamma_{ij}| \quad \text{par rapport aux variables } \Gamma_{ij} \quad (\xi_{ij} \in \mathbb{R}).$$

TAB. 3.1 – Propriétés souhaitées sur la partition, solution du modèle de classification

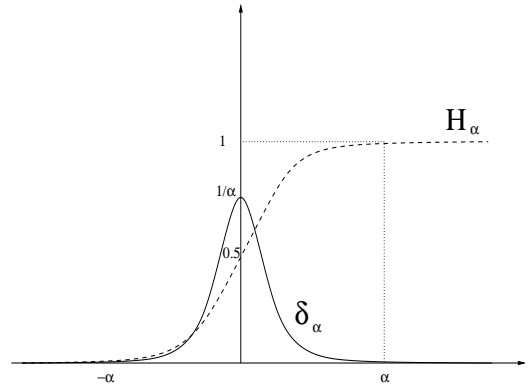
3.3.1 Préliminaires

Soit $\Phi_i : \Omega \rightarrow \mathbb{R}$ une fonction lipschitzienne continue³ associée à chaque région Γ_i telle que :

$$\begin{cases} \Phi_i(x) > 0 & \text{si } x \in \Omega_i \\ \Phi_i(x) = 0 & \text{si } x \in \Gamma_i \\ \Phi_i(x) < 0 & \text{sinon .} \end{cases} \quad (3.18)$$

L'ensemble des régions Ω_i ainsi que l'ensemble des frontières associées Γ_i sont entièrement décrites par la fonction Φ_i .

³on suppose l'existence d'une telle fonction pour toute région Γ_i

FIG. 3.4 – Approximations δ_α et H_α des distributions de Dirac et de Heaviside.

Remarque : Par la suite, nous omettrons la variable $x \in \Omega$ dans l'expression $\Phi_i(x)$ lorsque cela ne prête pas à confusion.

Nous définissons les approximations des distributions de Dirac δ_α et de Heaviside H_α (cf. figure 3.4), avec $\alpha \in \mathbb{R}^+$, par :

$$\delta_\alpha(s) = \begin{cases} \frac{1}{2\alpha} \left(1 + \cos\left(\frac{\pi s}{\alpha}\right) \right) & \text{si } |s| \leq \alpha \\ 0 & \text{si } |s| > \alpha \end{cases} \quad (3.19)$$

$$H_\alpha(s) = \begin{cases} \frac{1}{2} \left(1 + \frac{s}{\alpha} + \frac{1}{\pi} \sin\left(\frac{\pi s}{\alpha}\right) \right) & \text{si } |s| \leq \alpha \\ 1 & \text{si } s > \alpha \\ 0 & \text{si } s < -\alpha \end{cases} \quad (3.20)$$

et nous avons :

$$\begin{cases} \delta_\alpha \xrightarrow{\mathcal{D}'(\Omega)} \delta & \text{lorsque } \alpha \rightarrow 0^+ \\ H_\alpha \xrightarrow{\mathcal{D}'(\Omega)} H & \text{lorsque } \alpha \rightarrow 0^+ \end{cases}$$

où $\mathcal{D}'(\Omega)$ est l'espace des distributions défini sur Ω . D'après (3.18), (3.19) et (3.20) on obtient :

$$\{x \in \Omega / \lim_{\alpha \rightarrow 0^+} H_\alpha[\Phi_i(x)] = 1\} = \Omega_i \quad (3.21)$$

$$\{x \in \Omega / \lim_{\alpha \rightarrow 0^+} \delta_\alpha[\Phi_i(x)] \neq 0\} = \Gamma_i. \quad (3.22)$$

Les K ensembles Ω_i et Γ_i sont ainsi modélisés à l'aide des fonctions d'ensembles de niveaux Φ_i (cf. figure 3.5). Nous allons ainsi pouvoir traduire les contraintes sur la solution de

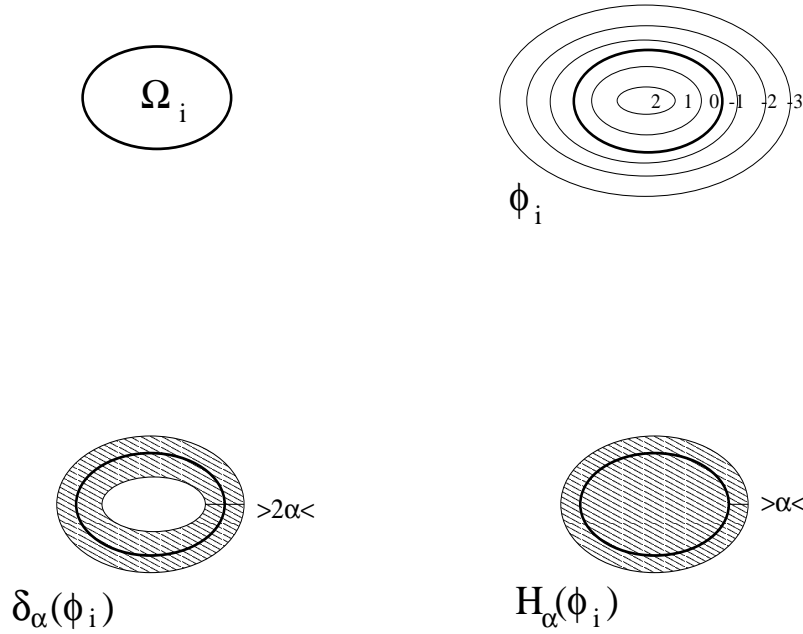


FIG. 3.5 – A chaque ensemble de régions Ω_i est associée une fonction de distance signée Φ_i . Les régions hachurées représentent les zones non nulles lorsque l'on applique une approximation de la distribution de Dirac δ_α ou de Heaviside H_α sur la fonction Φ_i .

partition (exprimées dans le tableau 3.1) à travers une fonctionnelle définie sur l'ensemble des fonctions Φ_i .

3.3.2 Expression de la fonctionnelle

D'après la définition (3.18) des fonctions Φ_i et en utilisant les formules (3.21) et (3.22), nous allons trouver une partition $\{\Omega_i\}_{i=1\dots K}$ de Ω vérifiant les conditions A, B et C du tableau 3.1 en minimisant une fonctionnelle définie sur l'ensemble des fonctions Φ_i . Cette fonctionnelle contiendra trois termes, chacun d'entre eux étant la traduction d'une des trois conditions du tableau 3.1 dans le cadre d'une formulation par ensembles de niveaux. Dans ce qui suit, nous allons exprimer chacune de ces trois conditions comme résultant de la minimisation d'une fonctionnelle définie sur les fonctions d'ensembles de niveaux Φ_i .

- FONCTIONNELLE LIÉE À LA CONDITION A (CONDITION DE PARTITION) :

Considérons la fonctionnelle suivante :

$$F_\alpha^A(\Phi_1, \dots, \Phi_K) = \frac{\lambda}{2} \int_{\Omega} \left(\sum_{i=1}^K H_\alpha(\Phi_i) - 1 \right)^2 dx \quad \text{avec } \lambda \in \mathbb{R}^+. \quad (3.23)$$

La minimisation de F_α^A , lorsque $\alpha \rightarrow 0^+$, conduit à une solution dont les régions vides et le chevauchement de régions sont pénalisés. La minimisation de ce terme permet de contraindre les pixels de la solution de partition à n'avoir qu'une et une seule étiquette attribuée, autrement dit à maintenir la contrainte

$$\sum_{i=1}^K H_\alpha(\Phi_i(x)) = 1, \quad p.p \text{ sur } \Omega.$$

- FONCTIONNELLE LIÉE À LA CONDITION B (ATTACHE AUX DONNÉES) :

Prenant en compte les caractéristiques des classes, à savoir la répartition gaussienne de l'intensité I dans chaque classe, et compte tenu de la remarque formulée plus loin dans le paragraphe 3.6, nous définissons la fonctionnelle d'attache aux données suivante :

$$F_\alpha^B(\Phi_1, \dots, \Phi_K) = \sum_{i=1}^K e_i \int_{\Omega} H_\alpha(\Phi_i) \frac{(I(x) - \mu_i)^2}{\sigma_i^2} dx \quad \text{avec } e_i \in \mathbb{R}^+, \forall i \in [1, \dots, K]. \quad (3.24)$$

Les paramètres e_i peuvent être utiles pour prendre en compte un a priori spécifique à chaque classe comme par exemple une mauvaise estimation des paramètres. Lorsque $\alpha \rightarrow 0^+$, la famille $\{\Phi\}_{i=1 \dots K}$ minimisant F_α^B conduit à une partition satisfaisant la condition B du tableau 3.1.

Il est intéressant d'analyser l'évolution des Ω_i localement sans considérer de contrainte sur la longueur des interfaces. Pour cela, supposons qu'il n'y ait que deux classes. Alors une seule fonction Φ est nécessaire et la condition de partitionnement est inutile. Le critère s'écrit, si la région Ω_1 est définie par l'ensemble $\{x/\Phi(x) > 0\}$ (alors la région complémentaire $\Omega_2 = \Omega \setminus \overline{\Omega_1}$ est définie par l'ensemble $\{x/\Phi(x) < 0\}$)

$$\int_{\Omega} H_\alpha(\Phi) \left(\frac{I - \mu_1}{\sigma_1} \right)^2 + \int_{\Omega} (1 - H_\alpha(\Phi)) \left(\frac{I - \mu_2}{\sigma_2} \right)^2 \quad (3.25)$$

L'équation d'évolution dynamique de Φ , déduite de l'équation d'Euler de (3.25) est

$$\Phi^{t+1} = \Phi^t - dt \delta_\alpha(\Phi) \left[\frac{(I - \mu_1)^2}{\sigma_1^2} - \frac{(I - \mu_2)^2}{\sigma_2^2} \right]. \quad (3.26)$$

Ainsi, si l'on considère un point a tel que $\Phi(a) = 0$, et tel que $I(a) \sim \mathcal{N}(\mu_1, \sigma_1)$, i.e. a est dans la classe 1, comme sur la figure 3.6, en oubliant un moment le terme de régularisation

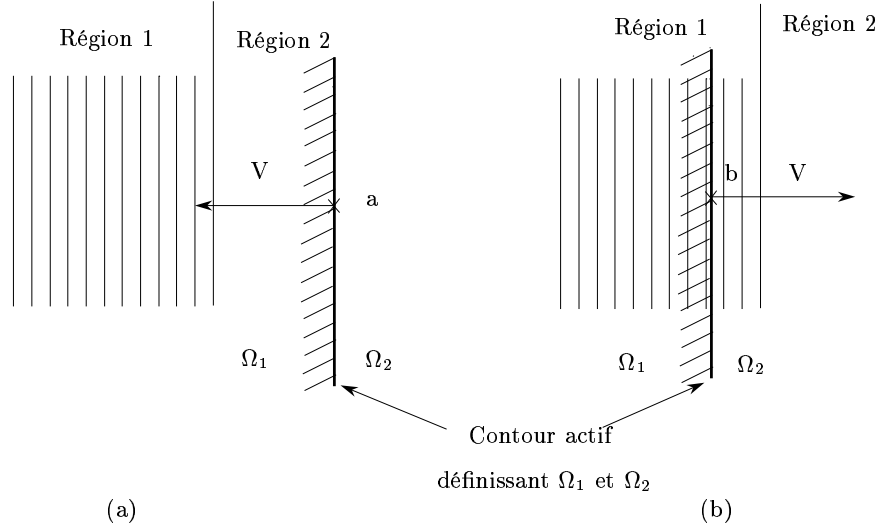


FIG. 3.6 – Etude locale de la vitesse du contour actif sur 2 régions. (a) Localement en a , $I(a) = \mu_2$ et la vitesse V est $(\mu_2 - \mu_1)^2 / \sigma_1^2$. (b) Localement en b , $I(b) = \mu_1$ et la vitesse V est $-(\mu_1 - \mu_2)^2 / \sigma_2^2$.

sur la longueur des contours actifs, alors la vitesse associée en a sera orientée comme sur la figure 3.6. Dès que le contour défini par $\Phi = 0$ arrive sur un point intérieur à la classe 2, alors le terme $\frac{(I-\mu_1)^2}{\sigma_1^2} - \frac{(I-\mu_2)^2}{\sigma_2^2}$ change de signe et la vitesse change donc de sens. Si le contour actif est sur Ω_1 , alors le terme vaut $-\frac{(\mu_1-\mu_2)^2}{\sigma_2^2}$, et si le contour actif est sur Ω_2 , alors ce terme vaut $\frac{(\mu_1-\mu_2)^2}{\sigma_1^2}$. Cette vitesse sera nulle lorsque le point a de l'ensemble de niveau 0 de Φ se trouve à la frontière entre les deux régions de classe 1 et 2, ainsi Ω_1 définit la classe 1 et Ω_2 définit la classe 2. Ce comportement est typique des contours actifs de *type région*, c'est-à-dire des contours actifs pour lesquels la vitesse est spécifiée par une donnée à l'intérieur (resp. l'extérieur) de la région, à la différence des contours actifs où la vitesse ne change pas de signe : le contour se dilate toujours ou se contracte toujours et s'arrête sur un potentiel nul (fonction $g(|\nabla I|)$).

• FONCTIONNELLE LIÉE À LA CONDITION C (MINIMISATION DE LA LONGUEUR DES INTERFACES) :

La dernière fonctionnelle que nous souhaitons définir est relative à la régularité morphologique de l'ensemble des contours actifs. Plus précisément, nous voulons minimiser :

$$\frac{1}{2} \sum_{i,j} \xi_{ij} |\Gamma_{ij}| \quad \text{avec } \xi_{ij} \in \mathbb{R}^+, \forall i, j \in [1, \dots, K]. \quad (3.27)$$

Le facteur $\frac{1}{2}$ est lié à la propriété de symétrie $\Gamma_{ij} = \Gamma_{ji}$ et sera inclus par la suite dans les valeurs de ξ_{ij} par souci de simplicité. Nous traduisons la minimisation de la longueur des interfaces en celle des frontières des classes :

$$\sum_{i=1}^K \gamma_i |\Gamma_i| \quad \text{avec } \gamma_i \in \mathbb{R}, \quad \forall i \in [1, \dots, K]. \quad (3.28)$$

A partir des équations (3.27) et (3.28) nous obtenons la relation $\xi_{ij} = \gamma_i + \gamma_j$ permettant de passer du problème de minimisation de la longueur des interfaces à celui de minimisation de la longueur des frontières. D'après le lemme qui suit, et à l'approximation δ_α près, minimiser (3.28) par rapport à l'ensemble $\{\Gamma_{ij}\}_{i,j}$ revient à minimiser par rapport à l'ensemble $\{\Phi\}_i$ la fonctionnelle suivante :

$$F_\alpha^c(\Phi_1, \dots, \Phi_K) = \sum_{i=1}^K \gamma_i \int_{\Omega} \delta_\alpha(\Phi_i) |\nabla \Phi_i| dx. \quad (3.29)$$

Lemme 1 : *Soit la fonctionnelle*

$$\mathcal{L}_\alpha(\Phi_i) = \int_{\Omega} \delta_\alpha(\Phi_i(x)) |\nabla \Phi_i(x)| dx,$$

alors nous avons

$$\lim_{\alpha \rightarrow 0} \mathcal{L}_\alpha(\Phi_i) = \int_{\Phi_i=0} ds = |\Gamma_i|.$$

La preuve de ce lemme, que nous donnons en annexe D (pour $g(|\nabla I|) = 1$), repose sur la formule de la co-aire.

• **FONCTIONNELLE GLOBALE :**

En introduisant les trois fonctionnelles F_α^A , F_α^B et F_α^C au sein d'une unique fonctionnelle, nous obtenons un critère global défini sur l'ensemble des fonctions Φ_i , et prenant en compte les trois conditions exprimées dans le tableau 3.1. La somme des trois fonctionnelles $F_\alpha^A + F_\alpha^B + F_\alpha^C$ nous permet de définir :

$$\begin{aligned} F_\alpha(\Phi_1, \dots, \Phi_K) = & \sum_{i=1}^K e_i \int_{\Omega} H_\alpha(\Phi_i) \frac{(I(x) - \mu_i)^2}{\sigma_i^2} dx + \sum_{i=1}^K \gamma_i \int_{\Omega} \delta_\alpha(\Phi_i) |\nabla \Phi_i| dx \\ & + \frac{\lambda}{2} \int_{\Omega} \left(\sum_{i=1}^K H_\alpha(\Phi_i) - 1 \right)^2 dx \end{aligned} \quad (3.30)$$

Lorsque $\alpha \rightarrow 0^+$, la famille $\{\Phi_i\}_{i=1\dots K}$ minimisant $F_\alpha(\Phi_1, \dots, \Phi_K)$, si elle existe, et compte tenu de la définition (3.18), définit une classification formulée par ensembles de niveaux

composée de régions formant des classes homogènes (les *phases* Ω_i) séparées par des interfaces régularisées.

3.3.3 Remarque concernant le problème de minimisation de longueur

Considérons la *fonctionnelle de longueur* suivante, où t représente le paramètre de temps :

$$L(t) = \int_0^1 \left| \frac{\partial C(p; t)}{\partial t} \right| dt \quad (3.31)$$

où $\{C(p; t)\}_t$ est un ensemble de courbes paramétrées fermées ($p \in [0; 1]$) définies sur Ω telles que $C(0; t) = C(1; t)$ et $\frac{\partial C(0; t)}{\partial t} = \frac{\partial C(1; t)}{\partial t}$. Alors, $L(t)$ décroît le plus rapidement pour l'évolution suivante :

$$\frac{\partial C(p; t)}{\partial t} = \kappa(p; t) \vec{N}, \quad (3.32)$$

où $\kappa(p; t)$ désigne la courbure de $C(p; t)$ et \vec{N} représentant la normale intérieur à $C(p; t)$. L'évolution d'une courbe selon (3.32) est connue sous le nom d'*évolution par courbure moyenne* (voir [Kichenassamy *et al.*, 1996] par exemple). L'EDP (3.32) peut être reformulée à l'aide d'une fonction d'ensemble de niveaux [Osher et Sethian, 1988], ce qui permet de gérer le changement de topologie de la courbe. Considérons la fonction lisse continue $d : \Omega \times \mathbb{R}^+ \rightarrow \mathbb{R}$ telle que $C(p; t) = \{x \in \Omega / d(x; t) = a\}$, c'est-à-dire que le contour C est modélisé à travers le niveau de valeur a de la fonction d . l'EDP (3.32) formulée par ensembles de niveaux devient :

$$\frac{\partial d(x; t)}{\partial t} = \operatorname{div} \left(\frac{\nabla d}{|\nabla d|} \right) |\nabla d|, \quad (3.33)$$

avec $\operatorname{div}(\frac{\nabla d}{|\nabla d|})$ représentant la courbure locale des niveaux de la fonction d . L'équation (3.33) a été étudiée par exemple par [Evans et Spruck, 1992]. L'évolution des niveaux de la fonction d (et par là même l'évolution de $C(p; t)$ à travers l'évolution du niveau de valeur a de d), selon l'EDP (3.33) est la formulation par ensembles de niveaux du mouvement par courbure moyenne du contour $C(p; t)$. Cette reformulation du mouvement autorise des brisures ainsi que des fusions de contours que ne permet pas la formulation de l'EDP (3.32). Dès lors que le contour $C(p; t)$ est représenté par le niveau a de d , il n'est pas nécessaire de considérer l'évolution de tous les niveaux de d , c'est pourquoi il est courant de n'appliquer l'EDP que pour les niveaux situés dans une bande étroite autour du niveau a (algorithme de type *narrow band* [Chopp, 1993, Malladi *et al.*, 1994]). La formulation par ensembles de niveaux de (3.33) provient d'une reformulation de (3.32) afin de gérer le changement de topologie lors de l'évolution du contour $C(p; t)$. Dans notre cas, nous utilisons une formulation par ensembles de niveaux dès le début, à savoir nous définissons la fonctionnelle de longueur

F_α^c sur les contours modélisés par les fonctions d'ensembles de niveaux Φ_i . Les équations d'Euler-Lagrange associées à la minimisation de F_α^c et plongées dans un schéma dynamique conduisent aux K EDP de la forme :

$$\frac{\partial \Phi_i(x; t)}{\partial t} = \operatorname{div} \left(\frac{\nabla \Phi_i}{|\nabla \Phi_i|} \right) \delta_\alpha(\Phi_i). \quad (3.34)$$

En comparaison du schéma (3.33), on voit apparaître dans (3.34) une bande étroite *naturelle* inhérente à la définition de l'approximation de Dirac δ_α , dont la largeur dépend de la valeur du paramètre α .

3.3.4 Fonction d'arrêt, information de type contour

Afin d'améliorer la convergence et la stabilisation des contours actifs au niveau des parois des objets à détecter, nous introduisons une fonction d'arrêt $g : \mathbb{R} \rightarrow \mathbb{R}^+$, comme dans [Caselles *et al.*, 1993], définie par :

$$g(|\nabla I|) = \frac{1}{1 + |\nabla I(x)|^2}.$$

Nous pouvons également définir

$$g(|\nabla I|) = \frac{1}{1 + |\nabla G_{\sigma_g} * I(x)|^2},$$

où G_σ est une Gaussienne de moyenne nulle et d'écart-type σ_g , et avec $*$ représentant le produit de convolution. Il s'agit donc de lisser les données avant de calculer le gradient, mais au risque de perdre de l'information sur les contours des objets (lissage isotrope), même si ce lissage présente un intérêt en présence de bruit. Dans nos expérimentations, nous ne filtrons pas les données par un lissage gaussien.

Nous introduisons la fonction g dans la fonctionnelle globale (3.30) pour obtenir la fonctionnelle suivante :

$$\begin{aligned} F_\alpha(\Phi_1, \dots, \Phi_K) = & \underbrace{\sum_{i=1}^K e_i \int_{\Omega} H_\alpha(\Phi_i) \frac{(I(x) - \mu_i)^2}{\sigma_i^2} dx}_{\text{attache aux données}} \\ & + \underbrace{\sum_{i=1}^K \gamma_i \int_{\Omega} g(|\nabla I|) \delta_\alpha(\Phi_i) |\nabla \Phi_i| dx}_{\text{minimisation de la longueur des contours}} \\ & + \underbrace{\frac{\lambda}{2} \int_{\Omega} \left(\sum_{i=1}^K H_\alpha(\Phi_i) - 1 \right)^2 dx}_{\text{contrainte de partition}}. \end{aligned} \quad (3.35)$$

L'introduction de la fonction g dans F_α induit un relâchement de la contrainte de régularité géométrique opérée sur les contours actifs lorsqu'ils rencontrent des zones de l'image où la valeur du module du gradient des données I est élevée⁴. On suppose que ces zones de l'image sont liées à la présence de frontières des objets à détecter.

3.4 Algorithme et schéma dynamique

En utilisant des conditions de Neumann aux bords de l'image pour les fonctions Φ_i :

$$\frac{\partial \Phi_i}{\partial \vec{n}}(x) = 0, \forall x \in \partial\Omega,$$

avec \vec{n} la normale unitaire à $\partial\Omega$, les équations d'Euler-Lagrange associées à la minimisation de (3.35) donnent les K EDP couplées suivantes :

$$\frac{\partial F_\alpha}{\partial \Phi_i} = 0, \forall i = 1 \dots K,$$

soit (les détails des calculs sont donnés en annexe D) :

$$\begin{aligned} \delta_\alpha(\Phi_i) \left[e_i \frac{(I - \mu_i)^2}{\sigma_i^2} - \gamma_i g(|\nabla I|) \operatorname{div} \left(\frac{\nabla \Phi_i}{|\nabla \Phi_i|} \right) - \gamma_i \frac{\nabla g \nabla \Phi_i}{|\nabla \Phi_i|} \right. \\ \left. + \lambda \left(\sum_{j=1}^K H_\alpha(\Phi_j) - 1 \right) \right] = 0, \quad \forall i = 1 \dots K \end{aligned} \quad (3.36)$$

avec $\operatorname{div} \left(\frac{\nabla \Phi_i}{|\nabla \Phi_i|} \right)$ la courbure moyenne de l'ensemble de niveaux Φ_i au point x . La présence du facteur $\delta_\alpha(\Phi_i)$ délimite une bande *naturelle* autour du niveau zéro de Φ_i dans laquelle la $i^{\text{ème}}$ EDP de (3.36) n'est pas nulle : $B_\alpha^i = \{x \in \Omega / |\Phi_i(x)| \leq \alpha\}$. Le terme $\frac{\nabla g \nabla \Phi_i}{|\nabla \Phi_i|}$ favorise la stabilisation des contours en présence de frontières d'objets dans les données I (cf. figure 3.1).

En plongeant (3.36) dans un schéma dynamique de pas de temps dt , on obtient un système de K équations couplées de la forme :

$$\begin{aligned} \Phi_i^{t+1} = \Phi_i^t - dt \left\{ \delta_\alpha(\Phi_i^t) \left[e_i \frac{(I - \mu_i)^2}{\sigma_i^2} - \gamma_i g(|\nabla I|) \operatorname{div} \left(\frac{\nabla \Phi_i^t}{|\nabla \Phi_i^t|} \right) - \gamma_i \frac{\nabla g \nabla \Phi_i^t}{|\nabla \Phi_i^t|} \right. \right. \\ \left. \left. + \lambda \left(\sum_{j=1}^K H_\alpha(\Phi_j^t) - 1 \right) \right] \right\}. \end{aligned} \quad (3.37)$$

avec $\Phi_i^t = \frac{\partial \Phi_i}{\partial t}$. Notons que nous initialisons les Φ_i par des fonctions de distances signées, qui sont le plus généralement utilisées dans les formulations par ensembles de niveaux pour

⁴on peut aussi bien considérer les valeur de $|\nabla I|$ que celles de $|\nabla G_{\sigma_g} * I|$

```

0 - initialiser  $\Phi_i^0$  pour  $i = 1 \dots K$ 
1 - pour  $i = 1 \dots K$  mettre à jour les  $K$  EDP couplées de (3.37)
2 - toutes les  $k$  itérations de (3.37), régulariser les
   fonctions  $\Phi_i$  à l'aide de l'EDP (3.7)

```

TAB. 3.2 – Algorithme associé au schéma (3.37).

leurs "bonnes" propriétés (cf. paragraphe 3.1). Cependant, comme pour (3.5), les équations (3.37) ne permettent pas de maintenir la contrainte $|\nabla \Phi_i| = 1$. Il est donc nécessaire d'imposer cette contrainte au cours de l'évolution de (3.37), par exemple par l'intermédiaire d'une EDP proposée dans [Sussman *et al.*, 1994] et appliquée régulièrement sur les fonctions Φ_i (cf. EDP (3.7) pour $u = \Phi_i$). On obtient alors l'algorithme dynamique présenté dans le tableau 3.2.

L'expression discrète de (3.37) en un pixel (m, n) de la grille est donnée en annexe F.

3.5 Résultats expérimentaux

Nous présentons quelques résultats de simulation et des tests menés sur des données synthétiques ainsi que sur des données réelles.

Nous tenons à remercier le Dr. Dormont de l'hôpital de La Pitié Salpêtrière pour les données IRM du cerveau traitées sur la figure 3.20.

Les ensembles de niveaux zéro "ENZ" représentent les contours des régions actives.

3.5.1 A propos du choix des paramètres et de l'initialisation

Les différents paramètres du modèle sont ajustés par l'expérimentateur. Il s'agit de régler le poids accordé aux différents termes de la fonctionnelle que nous proposons. Ces poids sont choisis en fonction des données, selon la taille et la morphologie des objets de l'image (poids du terme de longueur des contours), selon le bruit dans l'image (poids du terme d'attache aux données). Comme l'algorithme que nous utilisons est plongé dans un schéma dynamique (cf. équation (3.37)), nous devons choisir un pas de temps dt que nous fixons en général à 0.2. Nous réinitialisons les fonctions Φ_i toutes les 5 itérations par 5 étapes de l'EDP (3.7). Le paramètre d'approximation des distributions de Dirac et de Heaviside est fixe, sa valeur est fixée à 3.0 dans toutes les applications suivantes.

Initialisation des ENZ

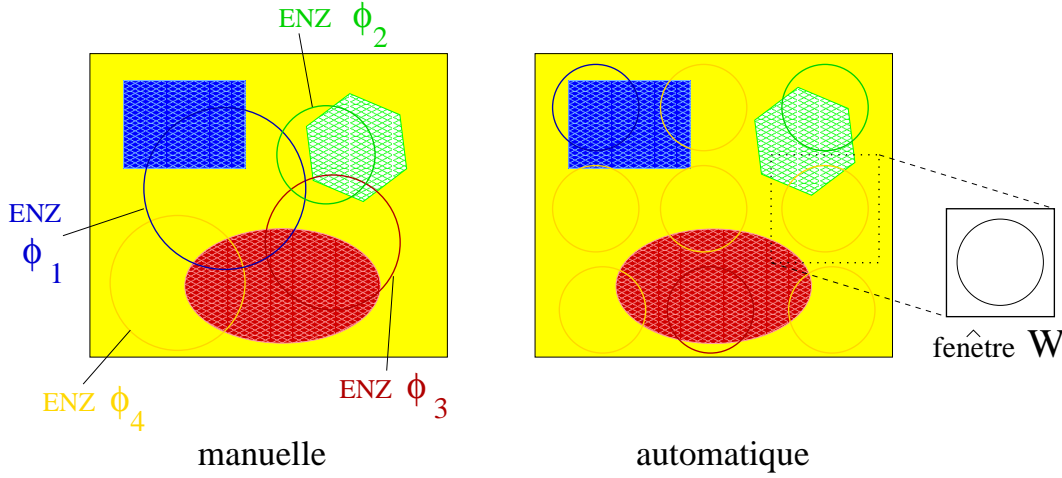


FIG. 3.7 – Les deux types d'initialisation des ENZ : manuelle (à gauche) ou automatique (à droite).

L'initialisation des fonctions Φ_i est opérée manuellement ou de façon automatique comme l'illustre la figure 3.7. De manière générale, nous choisissons des ensembles de niveaux circulaires. A partir du cercle constituant l'ENZ d'une fonction Φ_i , nous propageons les autres niveaux de cette fonction de distance signée. Une initialisation correcte nécessite d'avoir au moins un pixel de la classe C_i à l'intérieur de la région délimitée par l'ENZ de la fonction Φ_i . Lorsque l'initialisation est manuelle, l'opérateur positionne lui-même les fonctions Φ_i sur l'image. Pour une initialisation automatique, le principe consiste à découper l'image I en N_W fenêtres $W_{p,p=1..N_W}$ de taille fixée. Dans chaque fenêtre W_p , nous calculons la moyenne m_p et l'écart-type σ_p des données I , puis nous cherchons la classe la plus proche. La classe la "plus proche" est déterminée en trouvant l'indice k tel que $k = \arg \min_j d_B(\mathcal{N}_{(m_p, \sigma_p)}, \mathcal{N}_{(\mu_j, \sigma_j)})$, où d_B est la distance de Bhattacharyya [Basseville, 1985, Bhattacharyya, 1967, Fukunaga, 1972, Kittler, 1975] mesurant la distance entre deux distributions gaussiennes $\mathcal{N}_{(\mu_a, \sigma_a)}$ et $\mathcal{N}_{(\mu_b, \sigma_b)}$ par la relation :

$$d_B(\mathcal{N}_{(\mu_a, \sigma_a)}, \mathcal{N}_{(\mu_b, \sigma_b)}) = \frac{(\mu_b - \mu_a)^2}{4(\sigma_a + \sigma_b)} + \frac{1}{2} \log \frac{|\sigma_a + \sigma_b|}{2\sqrt{|\sigma_a \sigma_b|}}.$$

Ensuite, nous initialisons la fonction de distance signée Φ_k sur chaque fenêtre W_p . Les fenêtres $W_{p,p=1..N_W}$ ne se chevauchent pas et ne contiennent qu'une et une seule fonction Φ_k , évitant un recouvrement initial des régions Ω_k définies par les fonctions Φ_k . La taille des

fenêtres W_p est liée à la résolution des objets que l'on peut détecter. Une fenêtre trop petite ne contient pas suffisamment d'échantillons de données pour pouvoir obtenir une estimation significative des paramètres m_p et σ_p . Les avantages de cette méthode d'initialisation par rapport à l'initialisation manuelle sont les suivants : elle est automatique (seule la taille des fenêtres W_p doit être fixée), elle accélère la vitesse de convergence car les contours des régions actives sont plus proches des contours des classes à détecter, elle permet d'être moins sensible au bruit à partir du moment où nous calculons des moyennes sur les données pour initialiser les régions actives sans lisser les frontières des objets à détecter.

Remarque : Dans tout ce qui suit, les résultats de classification sont représentés en fausses couleurs. Les pixels noirs sur ces résultats correspondent à des pixels non classés (zones de vide). Les contours montrés sur les différentes figures, visualisent les frontières *internes* des régions actives. Ainsi, à convergence, la position réelle des interfaces entre les régions est située entre ces contours, d'où la présence de pixels sans étiquette au niveau des interfaces. On peut ensuite attribuer une étiquette à ces pixels de la façon suivante : à la fin de la convergence de l'algorithme, à chaque site s sans étiquette, on assigne l'étiquette $L_s \in \Lambda$ telle que :

$$L_s = \arg \max_{i=1..K} H_\alpha(\Phi_i(s)) \quad (3.38)$$

Lorsque les données synthétiques ne sont pas bruitées, nous prenons $\sigma_i = 1.0$, $\forall i = 1..K$.

3.5.2 Données synthétiques

Les figures 3.8 et 3.9 sont des simulations du modèle proposé. La figure 3.8 montre l'évolution de 3 contours de régions actives circulaires par courbure moyenne, c'est-à-dire lorsque les termes d'attache aux données ($e_i = 0$, $\forall i$) et de partition ($\lambda = 0$) de la fonctionnelle (3.35) sont nuls. L'évolution des fonctions Φ_i est donc guidée par le schéma dynamique :

$$\Phi_i^{t+1} = \Phi_i^t - dt \left\{ \delta_\alpha(\Phi_i^t) \left[\gamma_i g(|\nabla I|) \operatorname{div} \left(\frac{\nabla \Phi_i^t}{|\nabla \Phi_i^t|} \right) \right] \right\}.$$

Le poids du terme de contrainte de longueur γ_i est le même pour les trois classes. Les contours circulaires convergent vers des cercles dont le rayon diminue pour finir par un point et disparaître.

La figure 3.9 montre l'évolution des 3 contours de régions actives circulaires lorsque le terme d'attache aux données est nul. L'évolution des fonctions Φ_i est alors guidée par le schéma

dynamique :

$$\begin{aligned} \Phi_i^{t+1} = \Phi_i^t - dt \Big\{ \delta_\alpha(\Phi_i^t) \Big[-\gamma_i g(|\nabla I|) \operatorname{div} \left(\frac{\nabla \Phi_i^t}{|\nabla \Phi_i^t|} \right) - \gamma_i \frac{\nabla g \nabla \Phi_i^t}{|\nabla \Phi_i^t|} \\ + \lambda \left(\sum_{j=1}^K H_\alpha(\Phi_j^t) - 1 \right) \Big] \Big\}. \end{aligned}$$

Le poids de la contrainte de longueur est identique quelle que soit la classe. On remarque que les 3 cercles convergent vers les contours de 3 régions séparées par une jonction triple à 120 degrés [Merriman *et al.*, 1994, Reitich et Soner, 1996].

Les figures 3.10 et 3.11 montrent respectivement l'évolution des contours et des 3 régions circulaires pour des données constituées de 3 classes rectangulaires. Les valeurs des paramètres sont $\lambda = 5.0$, $dt = 0.2$, et pour tout $i = 1, 2, 3$ nous avons $\gamma_i = 1.0$ et $e_i = 2.0$. Le résultat de classification (en haut à droite de la figure 3.10) est en fausses couleurs.

Les figures 3.12 et 3.13 présentent respectivement l'évolution des contours et des 3 régions circulaires pour des données constituées de 3 classes de moyennes $\mu_1 = 0.0$, $\mu_2 = 1.0$ et $\mu_3 = 2.0$. Les valeurs des paramètres sont $\lambda = 2.0$, $dt = 0.2$, et pour tout $i = 1, 2, 3$ nous avons $\gamma_i = 0.4$ et $e_i = 2.0$. L'initialisation de la classe 2 est critique dans le sens où la classe 2 est constituée de 2 régions pour lesquelles nous n'avons initialisé qu'une région active.

Les figures 3.14 et 3.15 montrent respectivement l'évolution des contours et des 3 régions circulaires pour le bas des données "gdr" présentées dans les expérimentations du chapitre 2. Ces données sont constituées de 3 classes de moyennes $\mu_1 = 100.0$, $\mu_2 = 128.0$ et $\mu_3 = 160.0$. Les valeurs des paramètres sont $\lambda = 5.0$, $dt = 0.2$, et pour $i = 1, 2, 3$ nous avons $\gamma_i = 0.1$ et $e_i = 0.1$. L'intérêt de cette image réside dans les différentes courbures des objets la constituant. Le poids du terme d'attache aux données est faible compte tenu de l'importance de la valeur des moyennes des classes. La valeur des paramètres γ_i est également faible pour permettre aux contours actifs de détecter la frontière très oscillante de la forme située à droite de l'image "gdr".

Les données de la figure 3.16 sont celles de la figure 3.14 auxquelles nous avons ajouté un bruit gaussien de moyenne nulle (RSB=3.2dB). Nous avons initialisé les fonctions Φ_i par la méthode automatique, générant de petites régions circulaires sur des fenêtres de taille 5×5 pixels. Nous montrons l'évolution de contours et des régions sur la figure 3.16). Il y a un compromis à trouver concernant le poids accordé au terme de longueur des contours afin de ne pas être trop sensible au bruit (augmentation de la valeur des paramètres γ_i pour éviter la détection des régions de petite taille), tout en permettant aux contours actifs de segmenter la frontière à forte courbure de l'objet de droite (faible valeur de γ_3). La figure 3.17 montre l'influence de la fonction d'arrêt g dans (3.35) sur le résultat de classification

pour l'image "gdr" bruitée. La qualité des résultats est déterminée par rapport au nombre de pixels mal classés. On constate l'intérêt de l'utilisation de la fonction d'arrêt permettant d'introduire une information de type *contour* dans le modèle de régions actives proposé.

3.5.3 Données réelles

Nous avons conduit des tests sur des données médicales IRM ainsi que des images satellitaires SPOT.

La figure 3.18 montre les résultats obtenus sur une image IRM de cerveau contenant 4 classes dont les paramètres des classes ont été estimés au préalable par un algorithme de type *K means* (cf. chapitre 1), ainsi que trois étapes de l'évolution des contours et régions actifs.

Les données satellitaires SPOT des figures 3.19 et 3.20 sont celles traitées au chapitre 2 et dont les paramètres respectifs sont présentés dans le tableau 2.3.

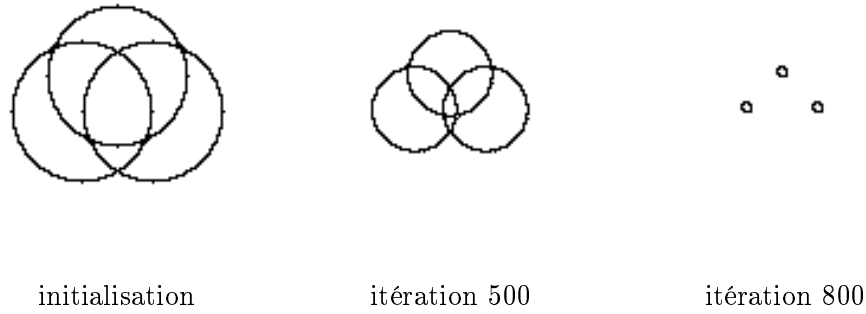


FIG. 3.8 – Evolution des trois ENZ par courbure moyenne. Les paramètres λ et e_i de (3.37) sont nuls, $\gamma_i = 10.0 \forall i$, et $dt = 0.2$.

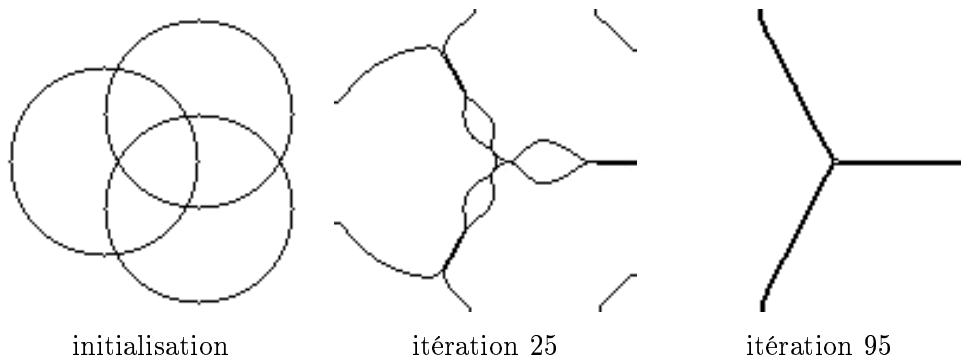


FIG. 3.9 – Evolution des trois ENZ sans terme d'attache aux données ($e_i = 0.0 \forall i$).

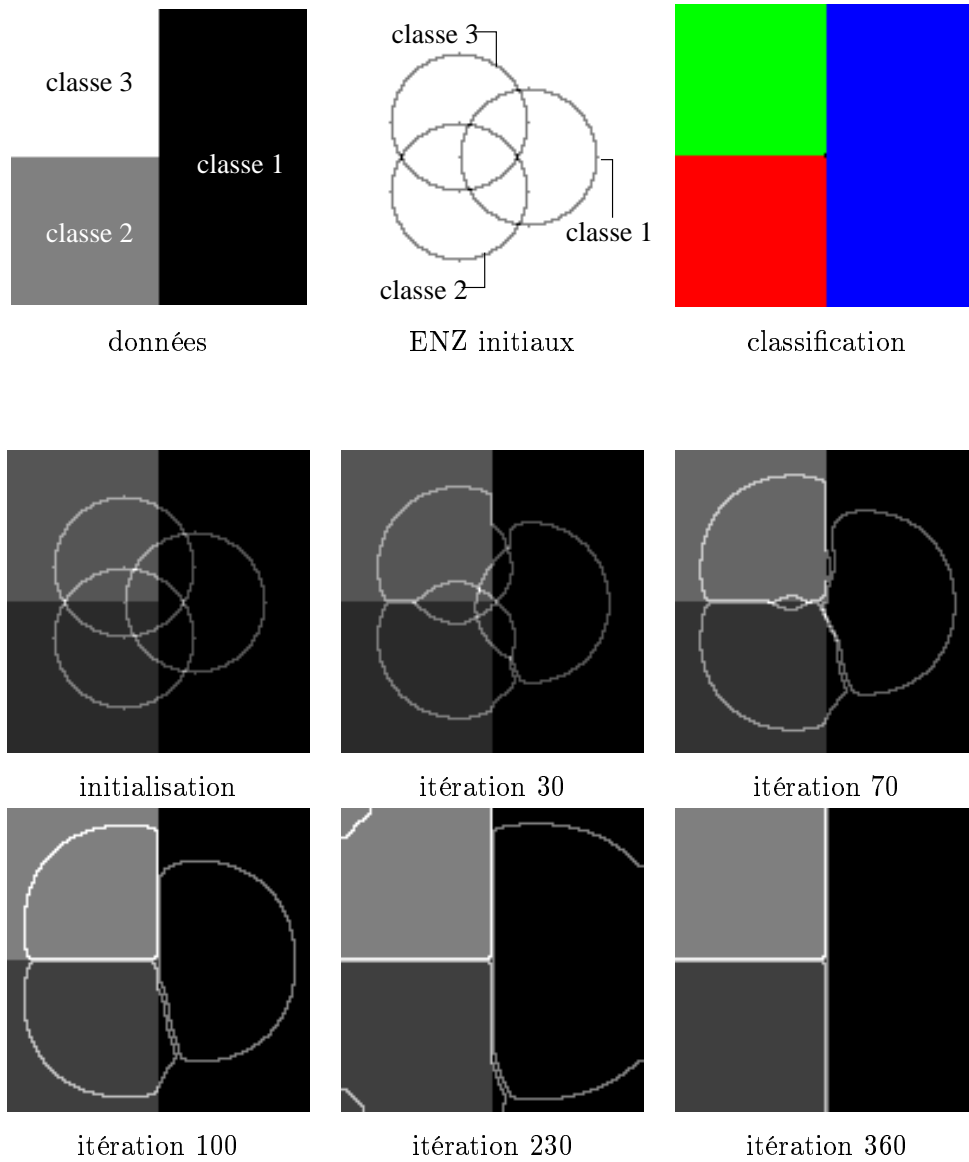


FIG. 3.10 – Evolution des ENZ pour les 3 classes de moyennes $\mu_1 = 1.0$, $\mu_2 = 2.0$ et $\mu_3 = 3.0$.

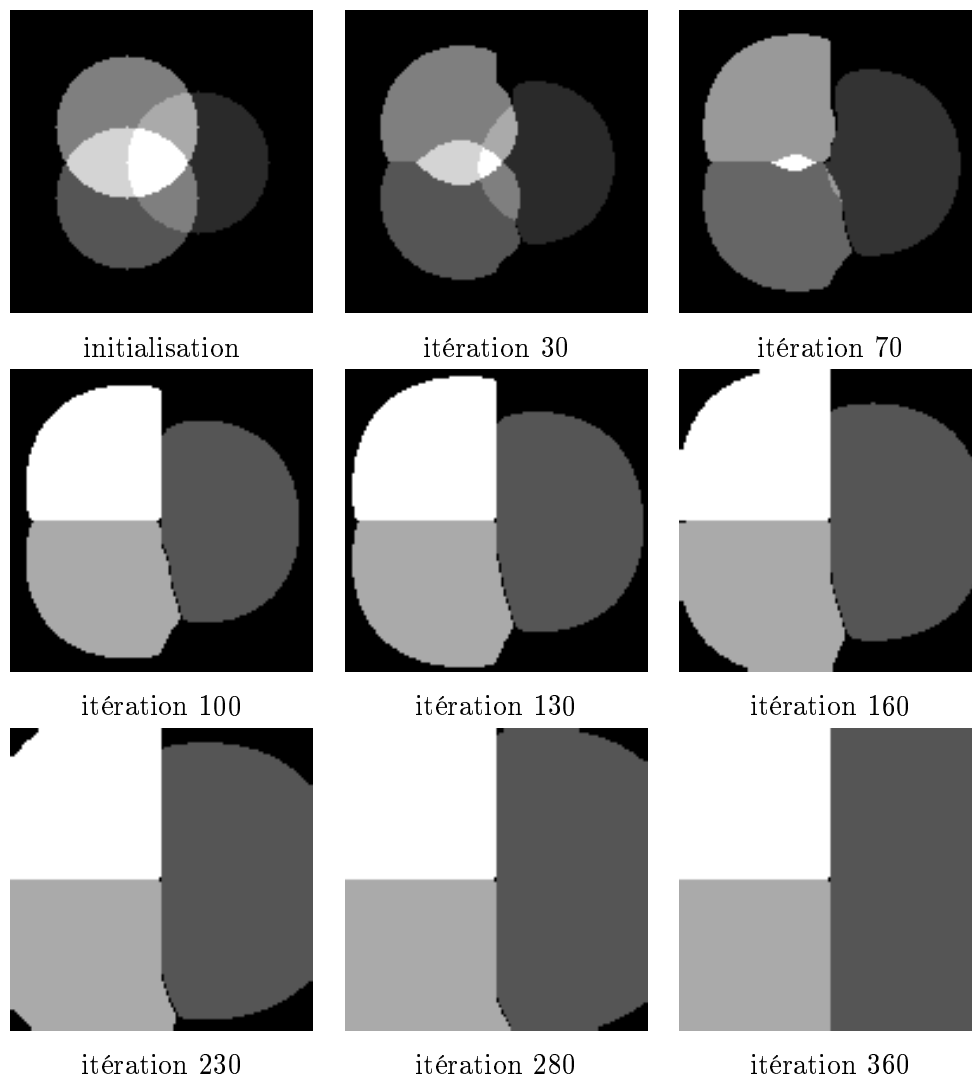


FIG. 3.11 – Evolution des régions associées à la figure 3.10.

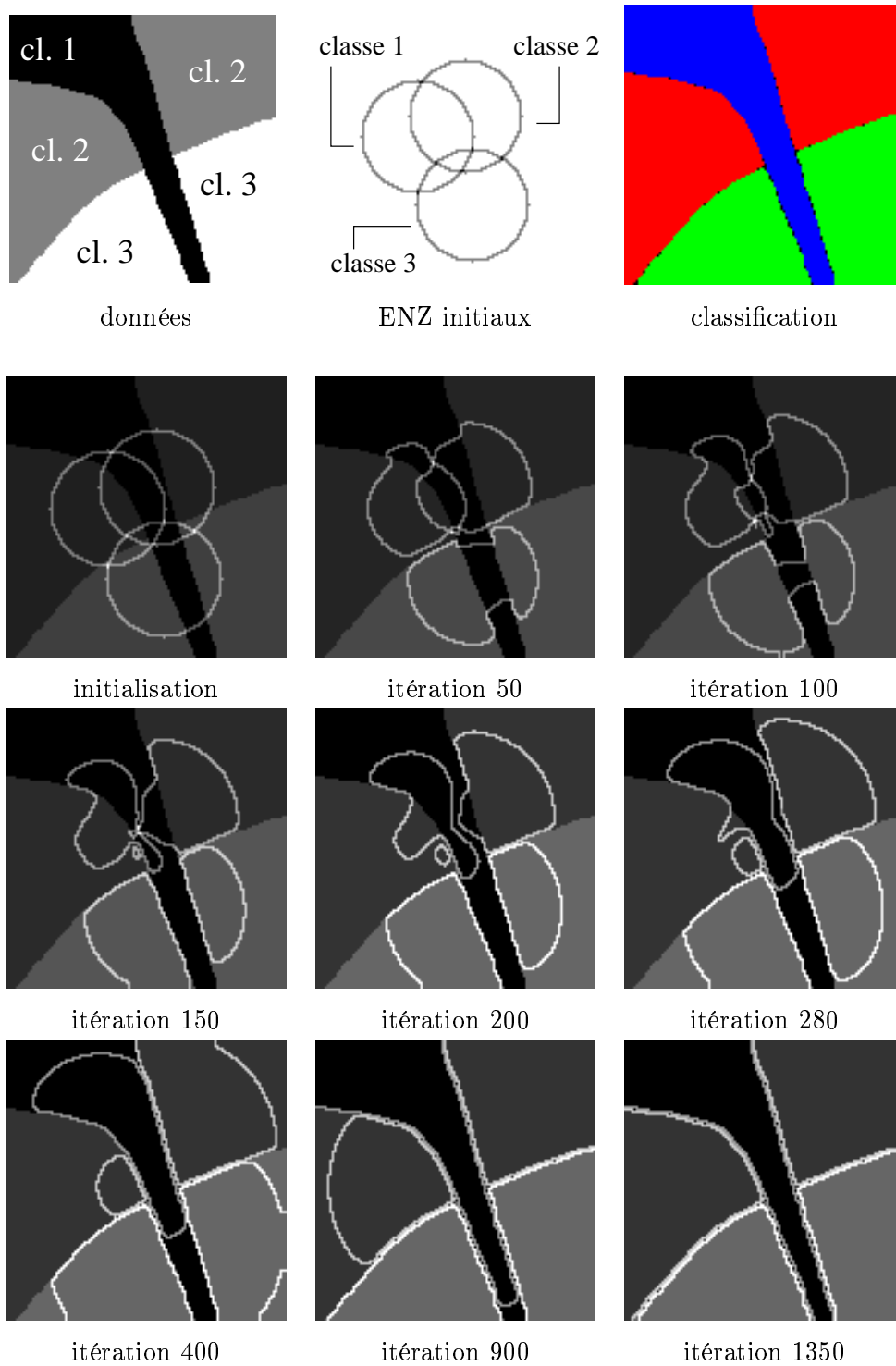


FIG. 3.12 – Evolution des ENZ pour les 3 classes.

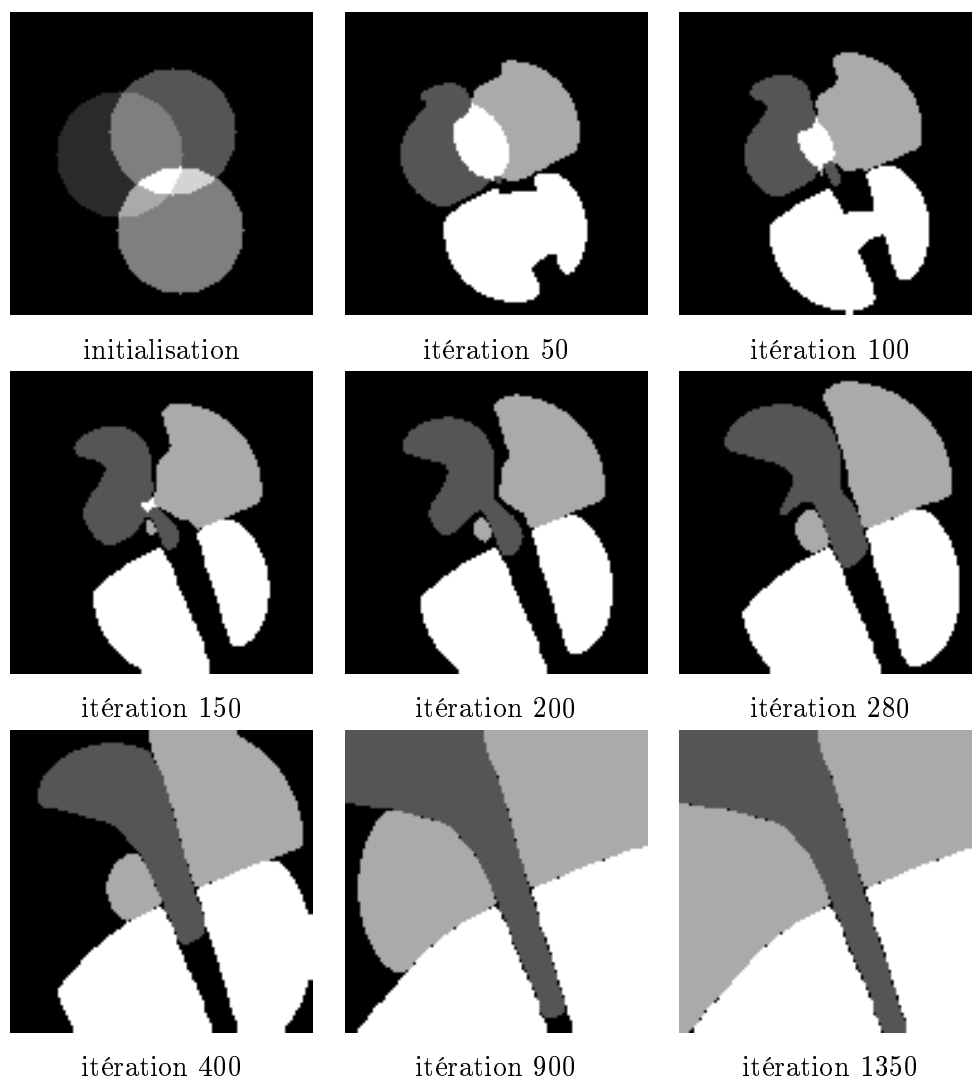


FIG. 3.13 – Evolution des régions associées à la figure 3.12.

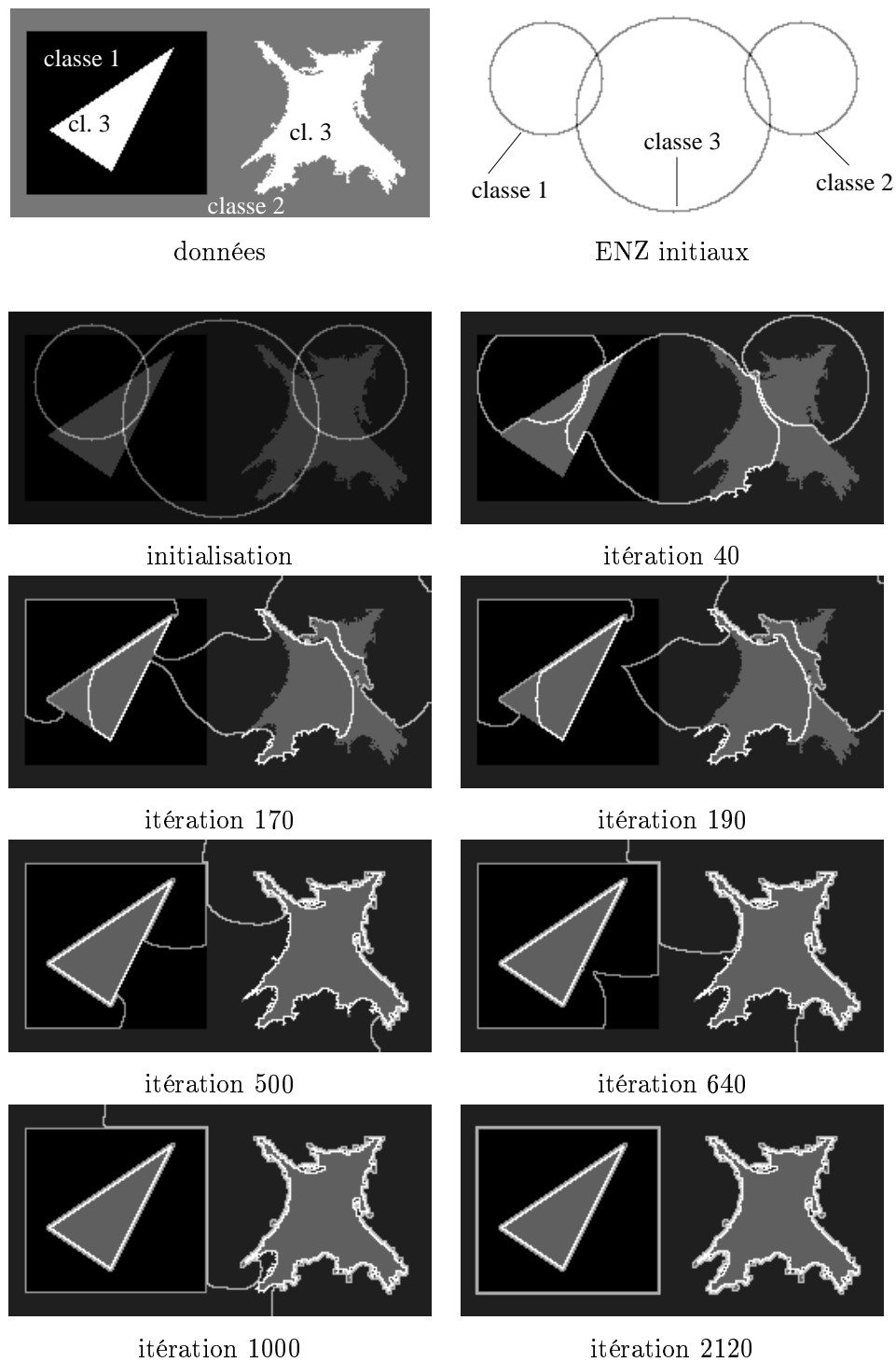


FIG. 3.14 – Image "gdr" : évolution des ENZ pour les 3 classes.

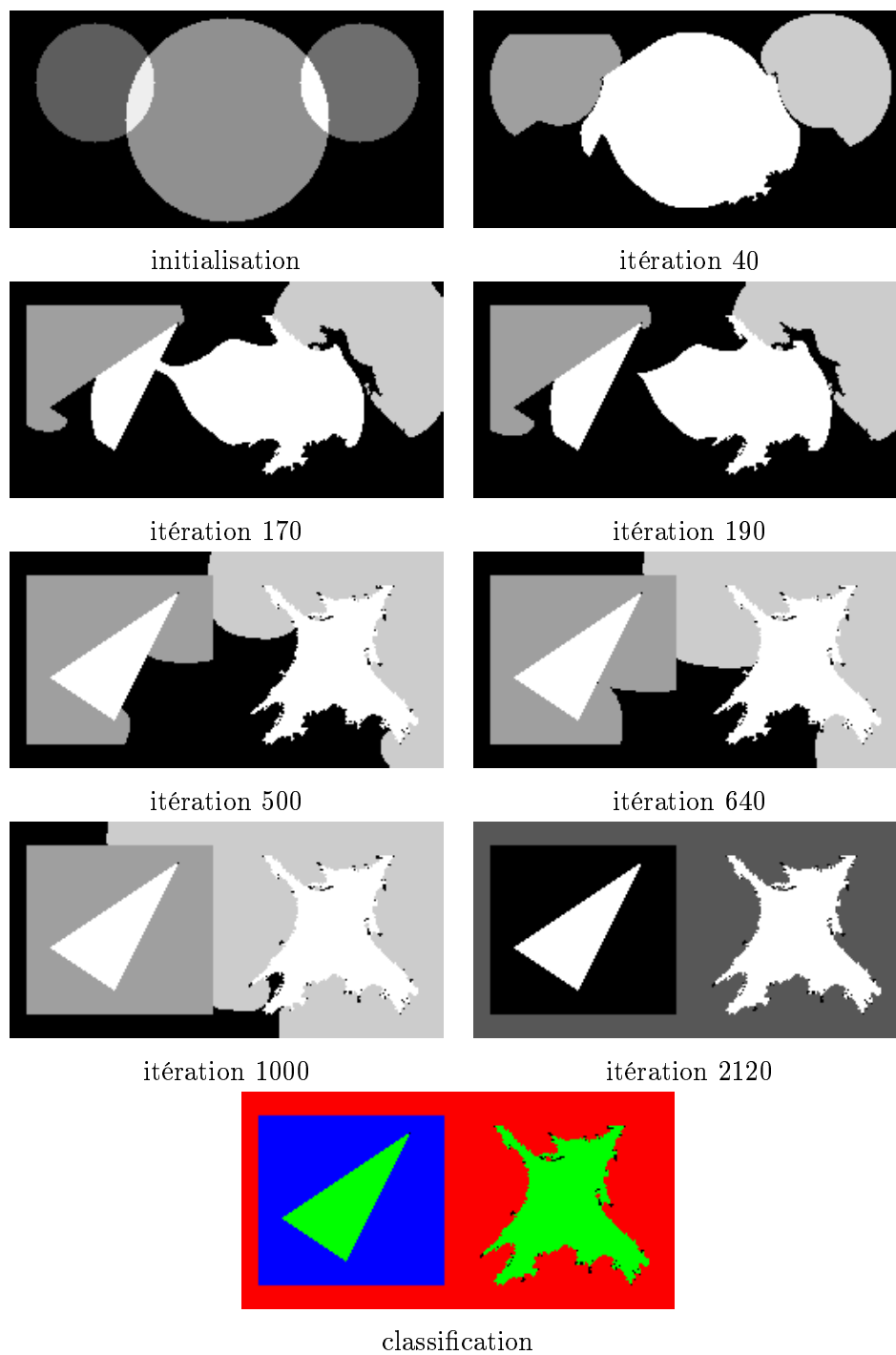


FIG. 3.15 – Image "gdr" : évolution des régions associées à la figure 3.14.

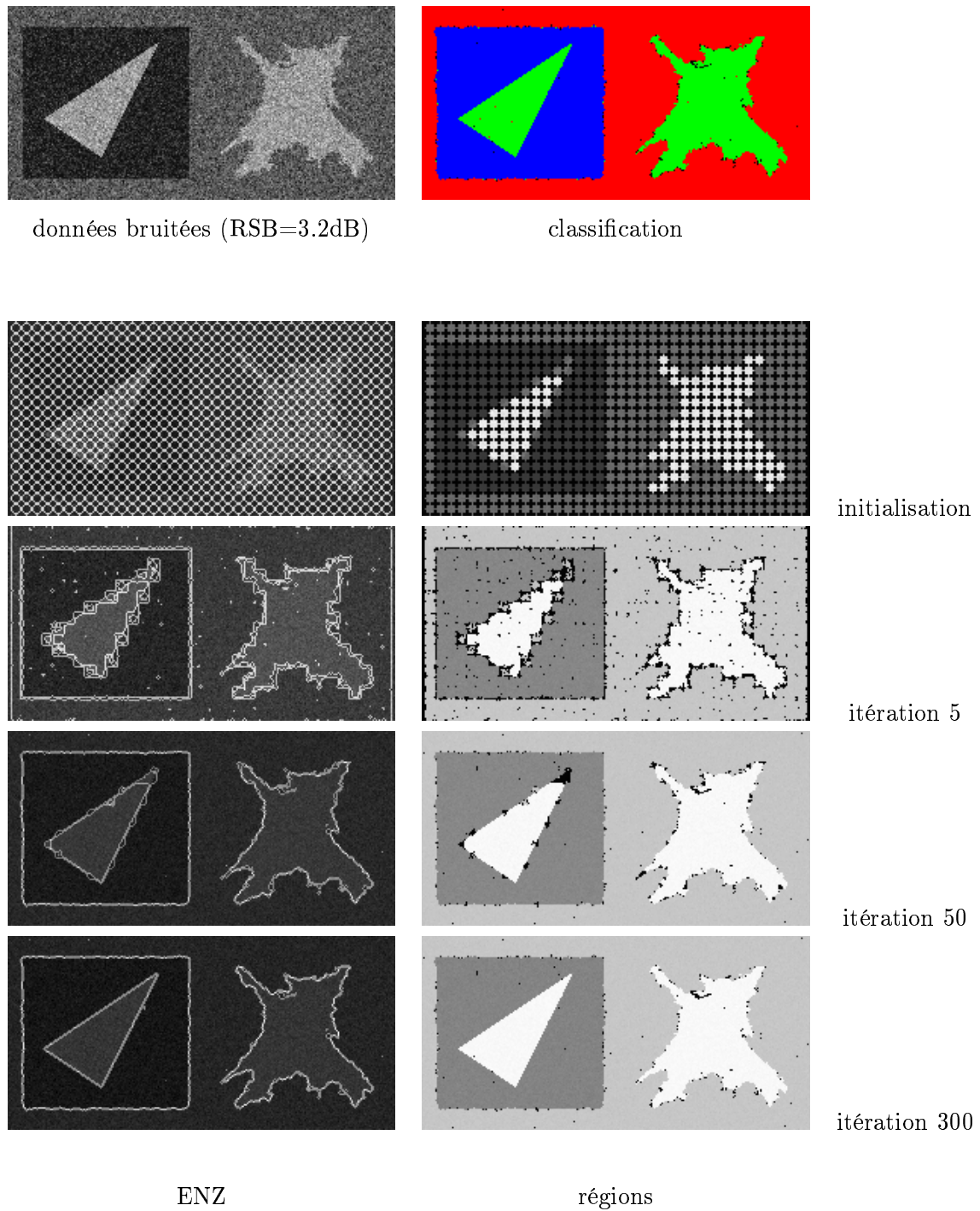


FIG. 3.16 – Image "gdr" : évolution des trois ENZ (colonne de gauche) et des régions associées (colonne de droite) sur la version bruitée (RSB=3.2dB) des données synthétiques présentées sur la figure 3.14.

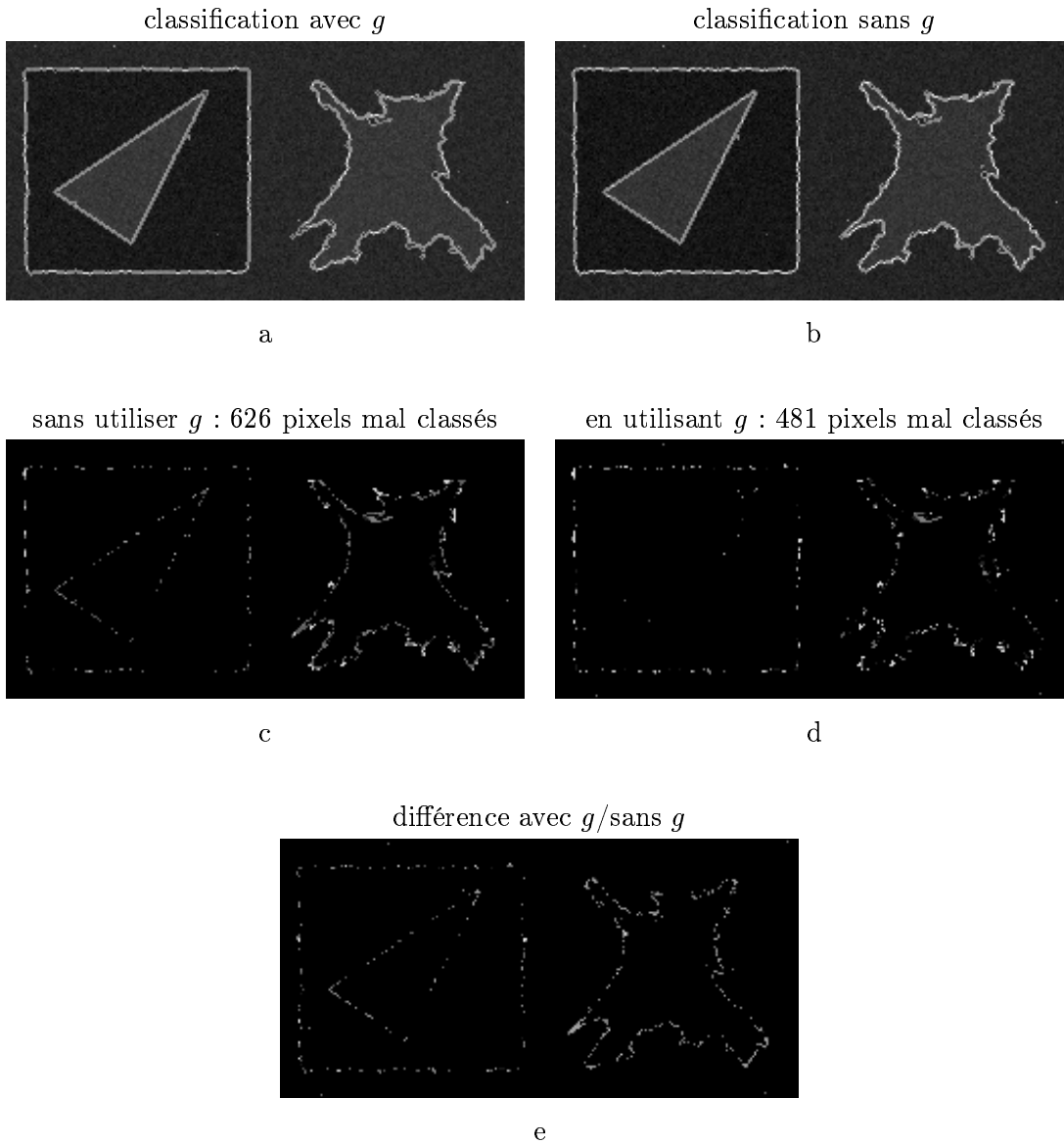


FIG. 3.17 – Influence de la fonction d'arrêt g sur la classification des données bruitées de la figure 3.16 en terme de pixels mal classés. Figure a : classification avec fonction d'arrêt g , figure b : classification sans fonction d'arrêt. Figure c : localisation des pixels mal classés lors de la classification sans fonction d'arrêt. Figure d : idem figure c pour une classification utilisant la fonction g . La figure e représente les pixels ayant obtenu une étiquette différente en utilisant ou pas g .

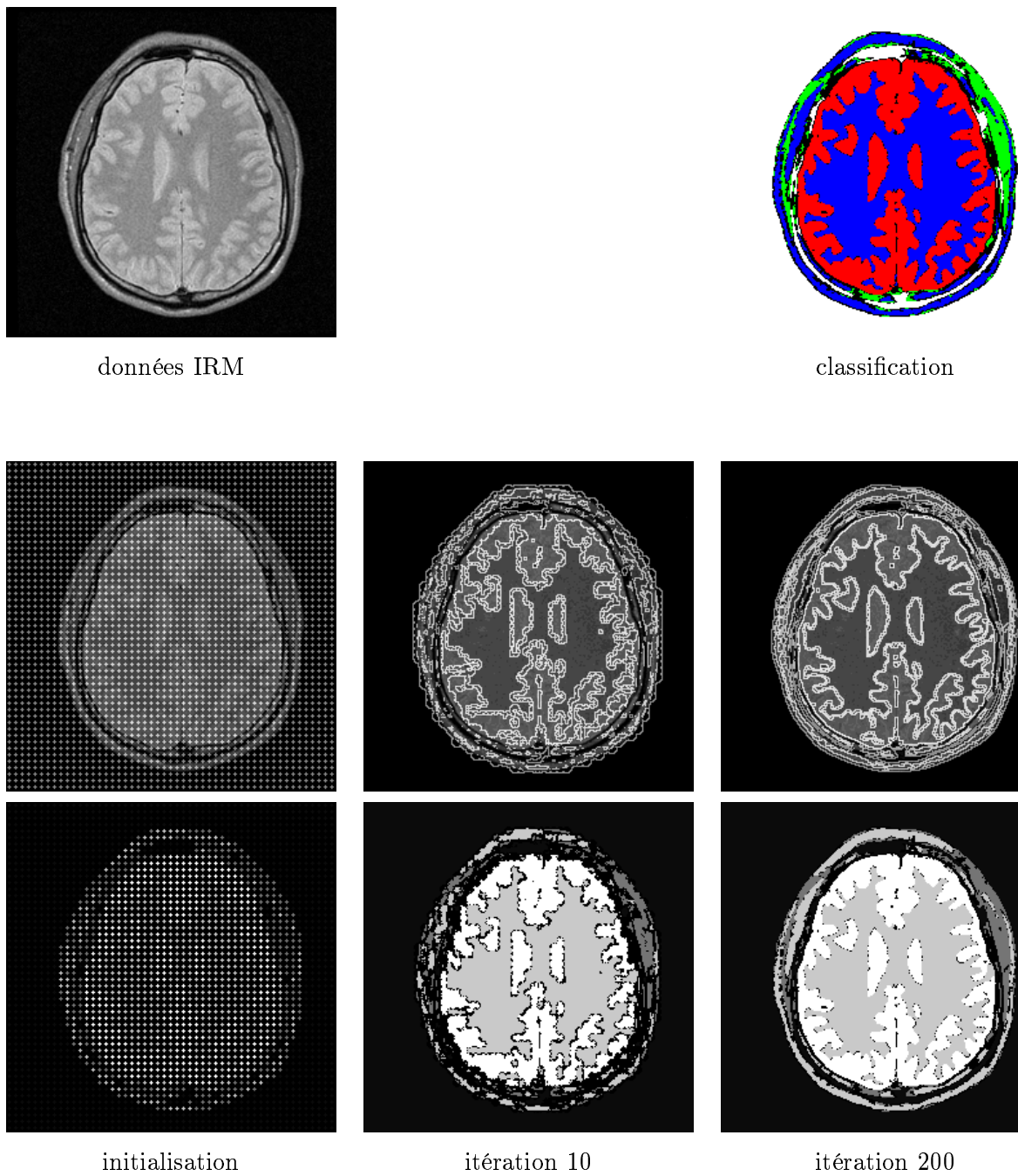
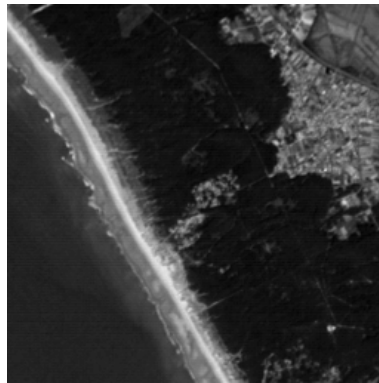
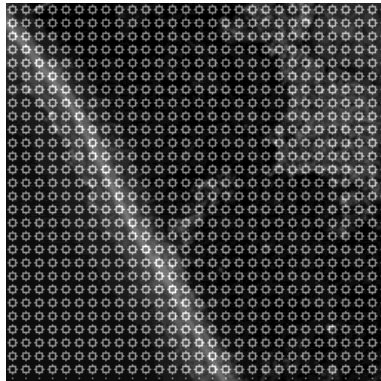


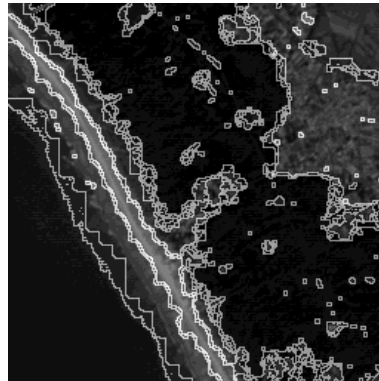
FIG. 3.18 – Image IRM du cerveau contenant 4 classes. Nous montrons 4 étapes de l'évolution des ENZ (en haut) et des régions associées (en bas).



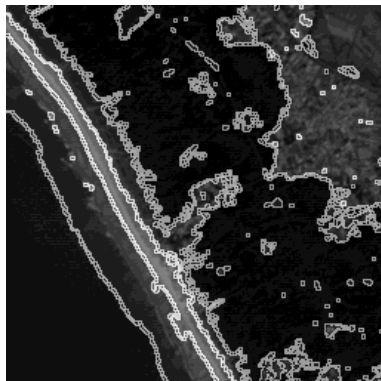
données SPOT



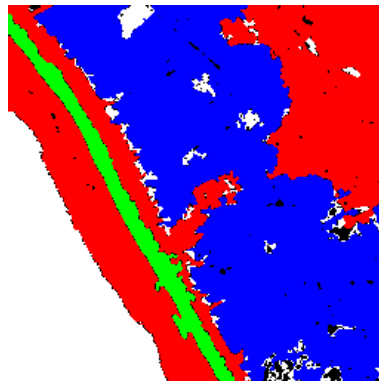
initialisation



itération 50



itération 300



classification

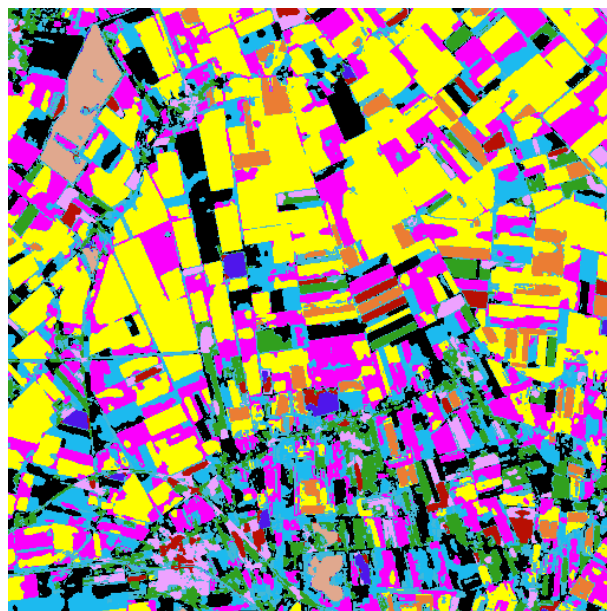
FIG. 3.19 – Données satellitaires SPOT (canal XS3) contenant 4 classes de paramètres pré-estimés. Nous montrons 3 étapes de l'évolution des ENZ. La dernière figure est le résultat de classification.



données SPOT



ENZ finaux



classification

FIG. 3.20 – Données satellitaires SPOT (canal XS3) contenant 10 classes de paramètres pré-estimés.

3.6 Remarques

3.6.1 Sur le terme d'attache aux données

Sur les termes globaux

Nous avons caractérisé une classe C_i par un ensemble de régions Ω_i sur lesquelles l'intensité I est distribuée de façon gaussienne :

$$\text{sur } \Omega_i, I(x) \sim \mathcal{N}(\mu_i, \sigma_i). \quad (3.40)$$

Le terme d'attache aux données de la fonctionnelle globale que nous avons proposé est le suivant (cf. (3.24)) :

$$F_\alpha^B(\Phi_1, \dots, \Phi_K) = \sum_{i=1}^K e_i \int_{\Omega} H_\alpha(\Phi_i) \frac{(I(x) - \mu_i)^2}{\sigma_i^2} dx \quad \text{avec } e_i \in \mathbb{R}, \forall i \in [1, \dots, K].$$

Il s'agit d'un terme local qui en tout point x de la région Ω_i retourne la distance euclidienne de l'observation $I(x)$ à la moyenne μ_i pondérée par la valeur de la variance correspondante. En notant respectivement m_i et v_i la moyenne et la variance des données I sur Ω_i , une traduction plus fidèle de la caractérisation des classes (3.40) peut s'écrire :

$$\text{sur } \Omega_i \begin{cases} m_i = \frac{1}{|\Omega_i|} \int_{\Omega} H_\alpha(\Phi_i) I dx = \mu_i \\ v_i = \frac{1}{|\Omega_i|} \int_{\Omega} H_\alpha(\Phi_i) [I - \frac{1}{|\Omega_i|} \int_{\Omega} H_\alpha(\Phi_i) I dx]^2 dx = \sigma_i^2 \end{cases}$$

avec $|\Omega_i|$ l'aire de Ω_i définie par $|\Omega_i| = \int_{\Omega} H_\alpha(\Phi_i) dx$. Le terme d'attache aux données correspondant à ces contraintes est le suivant :

$$\begin{aligned} F_\alpha^{\text{DONNÉES}}(\Phi_1, \dots, \Phi_K) = & \underbrace{\sum_{i=1}^K e_i \left[\frac{\int_{\Omega} H_\alpha(\Phi_i) I dx}{\int_{\Omega} H_\alpha(\Phi_i) dx} - \mu_i \right]^2}_{F_\alpha^{\text{MOY}}} \\ & + \underbrace{\sum_{i=1}^K e_i \left[\frac{1}{\int_{\Omega} H_\alpha(\Phi_i) dx} \int_{\Omega} H_\alpha dx \left[I - \frac{\int_{\Omega} H_\alpha(\Phi_i) I dx}{\int_{\Omega} H_\alpha(\Phi_i) dx} \right]^2 - \sigma_i^2 \right]^2}_{F_\alpha^{\text{VAR}}} \end{aligned}$$

Nous avons alors des termes globaux dans la fonctionnelle, représentés par les intégrations sur les domaines Ω_i , ce qui se traduit par une sommation sur l'ensemble des pixels au niveau discret. Ces termes globaux sont toujours délicats à traiter au niveau des EDP. Si nous ne considérons que le terme de pénalité sur la moyenne F_α^{MOY} , les K équations d'Euler-Lagrange sont de la forme :

$$\frac{\partial F_\alpha^{\text{MOY}}}{\partial \Phi_i} = 2e_i \delta_\alpha \frac{I \int_{\Omega} H_\alpha(\Phi_i) dx - \int_{\Omega} H_\alpha(\Phi_i) I dx}{[\int_{\Omega} H_\alpha dx]^2} \left[\frac{\int_{\Omega} H_\alpha(\Phi_i) I dx}{\int_{\Omega} H_\alpha(\Phi_i) dx} - \mu_i \right]. \quad (3.41)$$

Considérons alors les schémas dynamiques suivants pour $i = 1, \dots, K$:

$$\begin{aligned} \Phi_i^{t+1} = & \Phi_i^t - dt \left\{ \delta_\alpha(\Phi_i^t) \left\{ 2e_i \frac{I \int_\Omega H_\alpha(\Phi_i) dx - \int_\Omega H_\alpha(\Phi_i) I dx}{[\int_\Omega H_\alpha dx]^2} \left[\frac{\int_\Omega H_\alpha(\Phi_i) I dx}{\int_\Omega H_\alpha(\Phi_i) dx} - \mu_i \right] \right. \right. \\ & - \gamma_i g(|\nabla I|) \operatorname{div} \left(\frac{\nabla \Phi_i^t}{|\nabla \Phi_i^t|} \right) - \gamma_i \frac{\nabla g \nabla \Phi_i^t}{|\nabla \Phi_i^t|} \\ & \left. \left. + \lambda \left(\sum_{j=1}^K H_\alpha(\Phi_j^t) - 1 \right) \right\} \right\}. \end{aligned}$$

Quelques exemples permettent d'illustrer les problèmes liés aux termes globaux de l'attache aux données. Les données traitées sont celles présentées sur la figure 3.11 et contiennent 3 classes de moyennes $\mu_1 = 1.0, \mu_2 = 2.0$ et $\mu_3 = 3.0$ ($\sigma_i = 1.0$, pour $i = 1, 2, 3$). Si nous initialisons les ENZ des 3 classes comme dans le cas de la configuration B de la figure 3.22, nous voyons que chaque région circulaire de la colonne centrale supérieure des ENZ chevauche les deux régions de moyennes $\mu_1 = 1.0$ et $\mu_3 = 3.0$ (cf. détails des ENZ B sur la figure 3.21). Il y a autant de pixels de la classe 1 et 3 dans ces régions, ainsi la moyenne m_i des données sur les régions Ω_i de la colonne centrale vaut $(\mu_1 + \mu_3)/2 = \mu_2$. L'étiquette attribuée sera donc celle de la classe 2. Il n'y a pas de raison pour que cette étiquette change au cours de l'évolution car on ne considère pas la valeur locale des données dans ces régions, mais la valeur moyenne. Il restera donc une bande de la classe 2 chevauchant les deux régions de classe 1 et 3 en haut, au milieu de l'image (cf. évolution des régions et des ENZ de la configuration B sur la figure 3.22). En revanche, si l'initialisation est telle qu'il n'y a pas de chevauchement entre ces deux régions (cf. positionnement des ENZ A sur la figure 3.21 au niveau de la jonction triple), la classification est correcte : l'évolution des régions et des ENZ de la configuration A se comporte parfaitement comme l'illustre la figure 3.22.

La manipulation de termes globaux dans l'attache aux données conduit donc à des difficultés en cas de mauvaise initialisation et pour plus de deux classes.

Remarque

Nous avons défini la classe C_i par la distribution gaussienne :

$$\text{sur } \Omega_i, I(x) \sim \mathcal{N}(\mu_i, \sigma_i),$$

signifiant que la distribution de I sur Ω_i obéit à la loi :

$$\frac{1}{\sqrt{2\pi}\sigma_i} \exp \left[-\frac{(I(x) - \mu_i)^2}{2\sigma_i^2} \right].$$

Cette distribution exprime la vraisemblance des données introduite dans un cadre bayésien au sein du chapitre 1, formule (1.15), avec une hypothèse d'indépendance conditionnelle.

Dans le cadre d'une formulation par ensembles de niveaux, le terme d'attache aux données correspondant à cette distribution est :

$$F_{\alpha}^{\text{DONNÉES}}(\Phi_1, \dots, \Phi_K) = \sum_{i=1}^K e_i \int_{\Omega} H_{\alpha}(\Phi_i) \left[\frac{(I(x) - \mu_i)^2}{\sigma_i^2} + \log(\sqrt{2\pi}\sigma_i) \right] dx.$$

Nous avons donc négligé le terme $\sum_{i=1}^K e_i \int_{\Omega} H_{\alpha}(\Phi_i) \log(\sqrt{2\pi}\sigma_i) dx$ dans l'attache aux données F_{α}^{B} que nous utilisons pour le modèle dynamique de classification. Ce terme a peu d'influence lorsque les variances sont très proches d'une classe à l'autre, voire aucune influence lorsque toutes les classes ont même variance, ce qui est le cas pour les images synthétiques que nous avons traitées. En revanche, lorsque les valeurs des variances sont très différentes d'une classe à l'autre, ou encore lorsque toutes les classes ont même moyenne mais des variances différentes, il est nécessaire de prendre en compte ce terme.

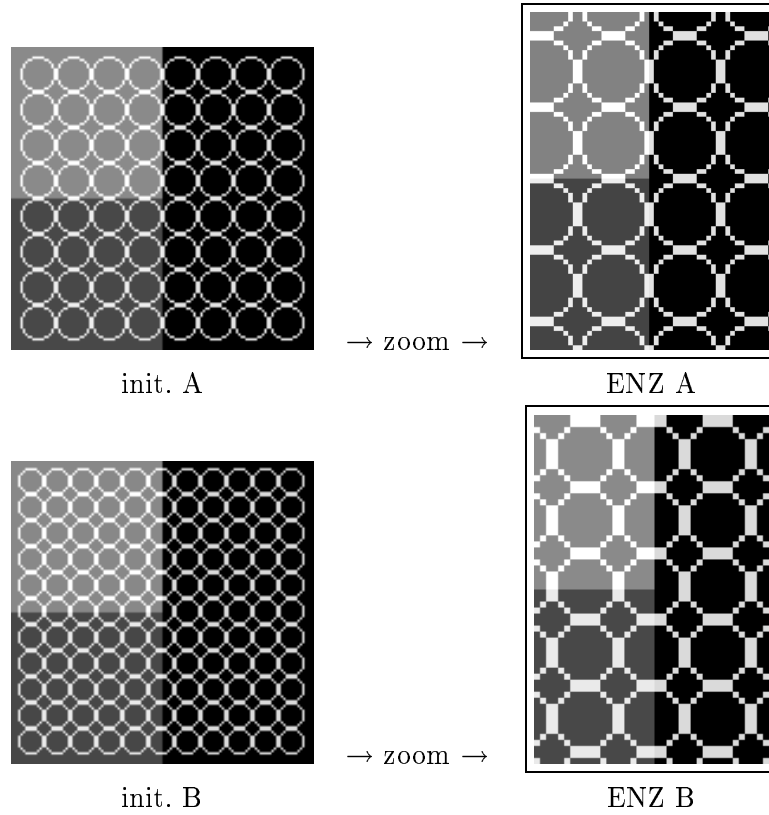


FIG. 3.21 – Initialisations des ENZ pour les configurations A et B de la figure 3.22, et zoom sur la position des ENZ au niveau de la jonction triple.

3.6.2 Sur l'introduction d'un processus de restauration

Nous proposons ici une façon d'introduire un processus de restauration avec lissage anisotrope préservant les contours dans la classification par modèle actif. Rappelons la fonctionnelle de classification (3.30) que nous avons défini :

$$F_\alpha(\Phi_1, \dots, \Phi_K) = \sum_{i=1}^K e_i \int_{\Omega} H_\alpha(\Phi_i) \frac{(I(x) - \mu_i)^2}{\sigma_i^2} dx + \sum_{i=1}^K \gamma_i \int_{\Omega} g(|\nabla I|) \delta_\alpha(\Phi_i) |\nabla \Phi_i| dx + \frac{\lambda}{2} \int_{\Omega} \left(\sum_{i=1}^K H_\alpha(\Phi_i) - 1 \right)^2 dx. \quad (3.42)$$

Supposons que I soit une version bruitée par un bruit gaussien de moyenne nulle et additif du signal original. Le seul moyen d'éviter de détecter de petites régions (et possédant généralement une forte courbure) dans I , dues à la présence de bruit, à l'aide de la minimisation de (3.42), est d'imposer un poids important au terme portant sur la longueur des contours (en jouant sur la valeur des paramètres γ_i). Dans la très grande majorité des cas, cela suffit pour traiter des données bruitées. Cependant, malgré la présence de la fonction d'arrêt g , permettant de relâcher la contrainte de longueur en présence de frontières d'objets lorsque le niveau de bruit est important, il est parfois délicat de trouver la "bonne" valeur des paramètres γ_i offrant la possibilité de détecter des objets de frontières oscillantes sans pour autant détecter les petites régions de forte courbure dues à la présence de bruit. Nous avons alors pensé à un moyen de combiner au modèle de classification un processus de lissage anisotrope. Pour cela, nous avons considéré un modèle de dégradation des données originelles comme dans l'équation (2.2) du chapitre 2 avec $\mathcal{R} = Id$ et f représentant la version non bruitée des données observées I . Nous avons défini la fonctionnelle suivante :

$$G_\alpha(f, \Phi_1, \dots, \Phi_K) = \sum_{i=1}^K \gamma_i \int_{\Omega} g(|\nabla f|) \delta_\alpha(\Phi_i) |\nabla \Phi_i| dx + \frac{\lambda_1}{2} \int_{\Omega} \left(\sum_{i=1}^K H_\alpha(\Phi_i) - 1 \right)^2 dx + \sum_{i=1}^K e_i \int_{\Omega} H_\alpha(\Phi_i) \frac{(f - \mu_i)^2}{\sigma_i^2} dx + \tau \left[\int_{\Omega} (I - f)^2 + \lambda_2 \int_{\Omega} \varphi(|\nabla f|) \right], \quad (3.43)$$

où φ est une fonction régularisante du même type que celles définies dans le tableau 2.1 du chapitre 2. Le réel τ est fixé manuellement et contrôle le poids accordé au terme de lissage anisotrope sur les données. De même, λ_1 et λ_2 sont des réels fixés. La fonctionnelle G_α dépend des K fonctions d'ensembles de niveaux Φ_i et également de la fonction f . Pour optimiser G_α , nous pouvons opter pour des minimisations alternées, en minimisant

successivement par rapport à f , puis par rapport aux variables Φ_i . Les équations d'Euler-Lagrange donnent respectivement les $(K + 1)$ EDP suivantes :

$$\frac{\partial G_\alpha}{\partial f} = \sum_{i=1}^K e_i H_\alpha(\Phi_i) \frac{(f - \mu_i)}{\sigma_i} + \Lambda \left[(f - I) - \lambda_2 \operatorname{div} \left(\frac{\varphi'(|\nabla f|)}{2|\nabla f|} \nabla f \right) \right] = 0, \quad (3.44)$$

et pour $i = 1 \dots K$ nous avons les K EDP couplées

$$\begin{aligned} \frac{\partial G_\alpha}{\partial \Phi_i} = \delta_\alpha(\Phi_i) \left[e_i \frac{(f - \mu_i)^2}{\sigma_i^2} - \gamma_i g(|\nabla f|) \operatorname{div} \left(\frac{\nabla \Phi_i}{|\nabla \Phi_i|} \right) \right. \\ \left. - \gamma_i \frac{\nabla g \nabla \Phi_i}{|\nabla \Phi_i|} + \lambda_1 \left(\sum_{j=1}^K H_\alpha(\Phi_j) - 1 \right) \right] = 0. \end{aligned} \quad (3.45)$$

l'équation (3.44) est résolue selon le même principe que l'équation (2.17) du chapitre 2 (*régularisation semi-quadratique* [Charbonnier *et al.*, 1997, Geman et Reynolds, 1992] présentée au paragraphe 2.2). Les K EDP couplées (3.45) sont plongées dans un schéma dynamique. On alterne la résolution de (3.44) et de (3.45) pour chaque itération en temps. Les premiers résultats ont été prometteurs, mais nous avons choisi de nous pencher sur le problème de l'extension au cas multispectral présentée au chapitre 5 dans le temps qui nous était imparti. Un exemple de résultat obtenu en résolvant (3.45), est montré sur la figure 3.23. Sur cette figure, nous montrons les résultats de classification obtenus avec le terme de restauration (minimisation de (3.43)) et sans restauration (minimisation de (3.35)) sur des données synthétiques bruitées à - 6.4dB contenant 3 classes. Nous présentons également l'évolution des ENZ, dont l'étape d'initialisation, l'itération 100 et les ENZ finaux. Nous constatons que l'introduction du processus de restauration améliore sensiblement les résultats surtout en ce qui concerne les zones de vide au niveau des frontières des classes (pixels noirs sur les résultats de classification en fausses couleurs) ainsi que la régularité des frontières.

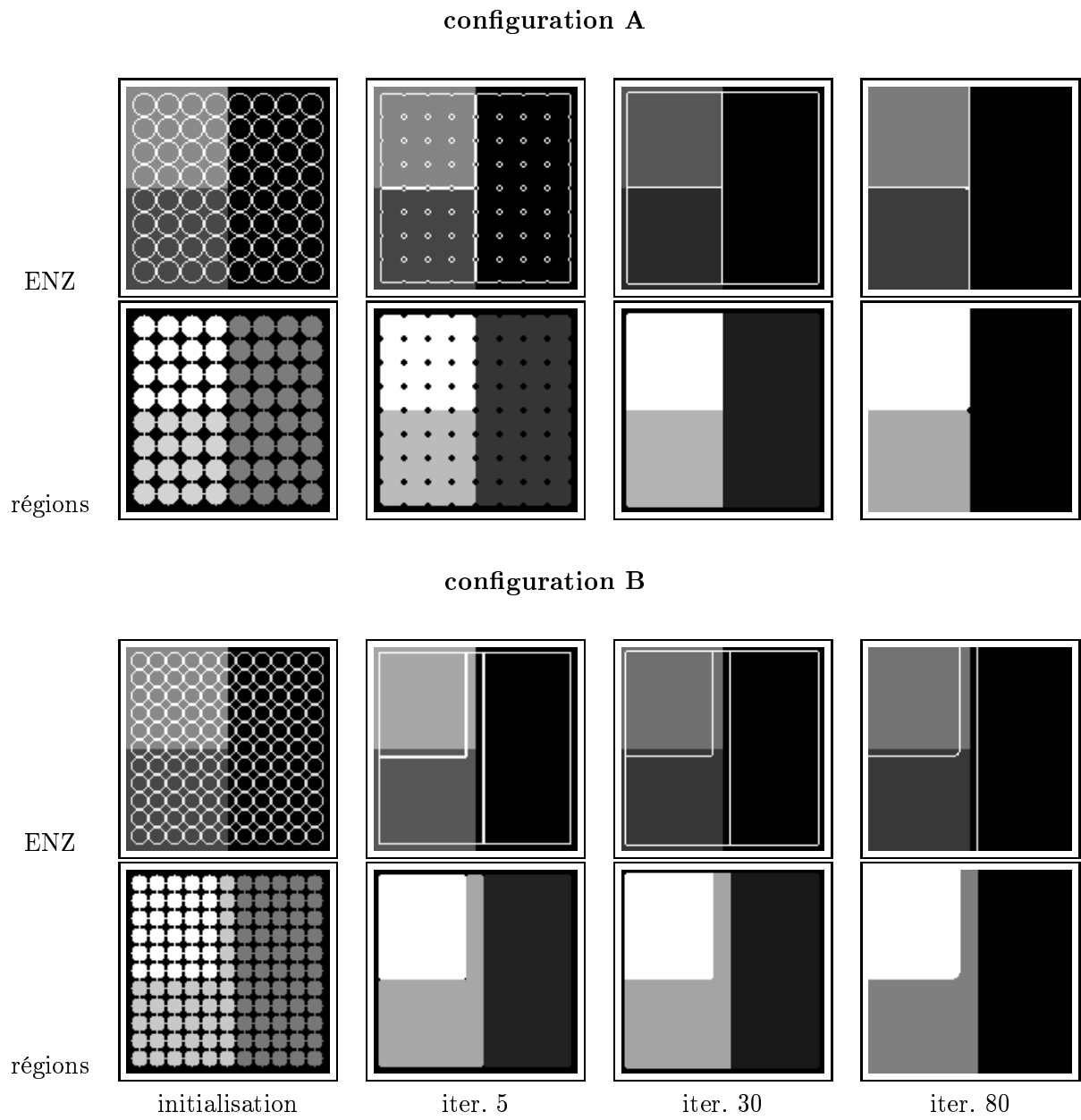


FIG. 3.22 – Deux types de configurations initiales : A et B.

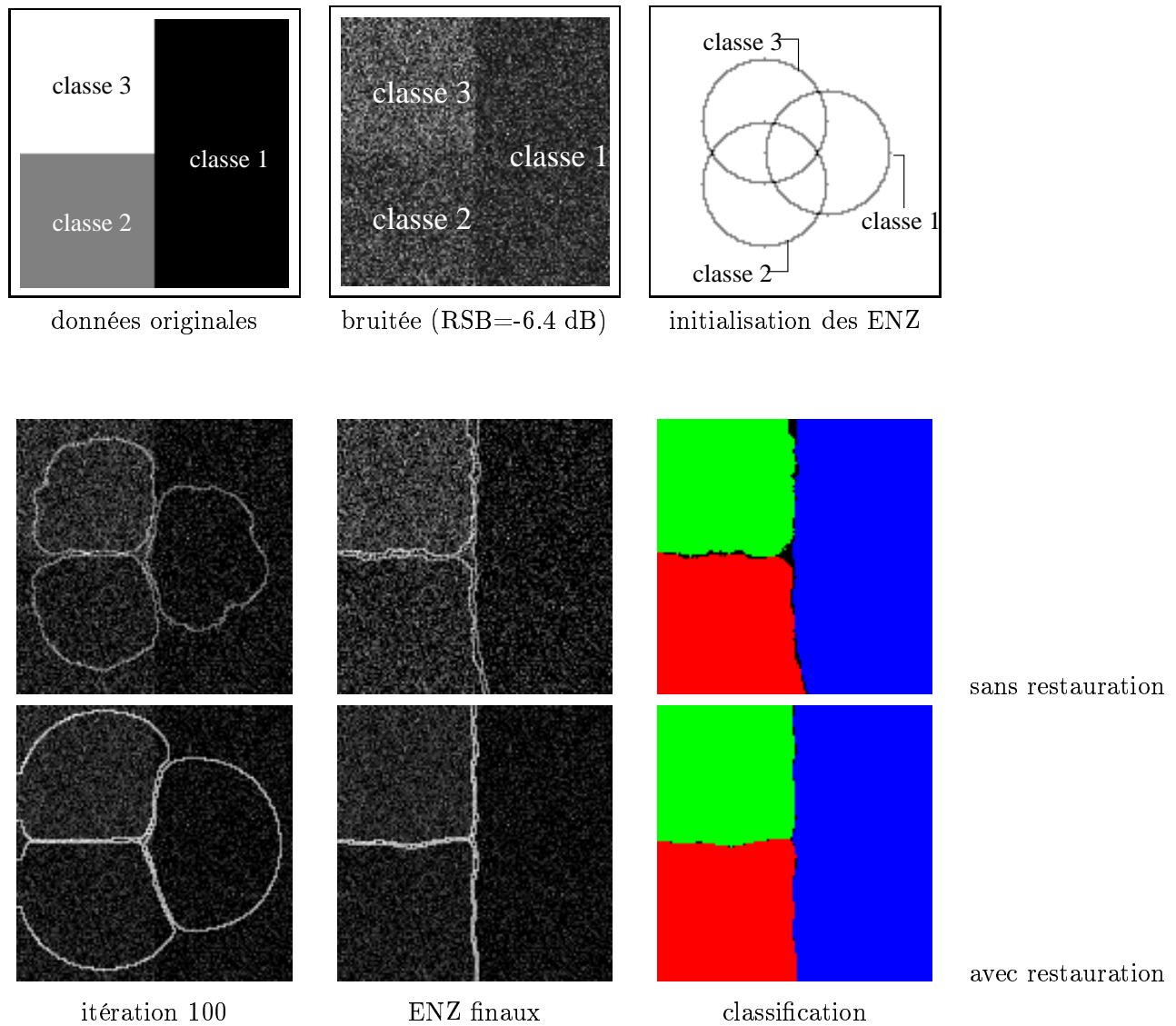


FIG. 3.23 – Influence du terme de restauration dans la classification par modèle actif.

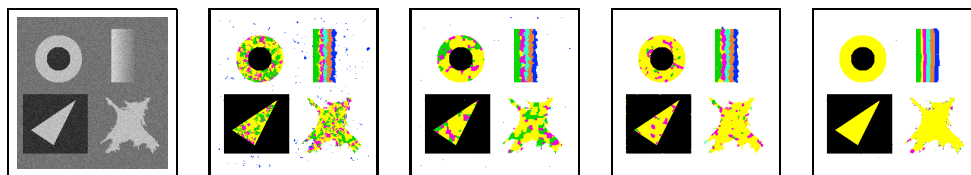
3.7 Conclusion

Nous avons présenté un modèle de classification mettant en jeu un ensemble de régions et de contours actifs. Ce modèle combine l'information contenue au niveau des régions ainsi que celle présente au niveau des frontières des objets de la scène observée.

Nous utilisons une formulation par ensembles de niveaux dont l'intérêt a été largement démontré dans le cadre de la segmentation d'images. Nous initialisons, de façon automatique ou manuelle, un ensemble de régions initiales pour chaque classe, qui vont évoluer par l'intermédiaire d'un schéma dynamique et converger vers une partition de l'image composée de régions homogènes dont les frontières sont régularisées. Remarquons que le nombre de classes n'est pas limité comme dans bon nombre de travaux menés dans ce cadre. La formulation par ensembles de niveaux, ainsi que les différentes contraintes sur la solution de classification sont directement exprimées au niveau de la fonctionnelle et non des EDP.

Chapitre 4

Quelques comparaisons



Où nous menons une étude comparative des résultats de classification obtenus à l'aide des deux modèles proposés dans les chapitre 2 et 3. Nous testons également les deux modèles stochastiques présentés dans le premier chapitre (le premier fondé sur une régularisation de Potts, le second sur le chien-modèle). Nous expérimentons ces quatre modèles sur des images synthétiques et réelles, et nous examinons le taux de réussite des différentes classifications en terme de pixels bien/mal classés, ainsi que le temps de calcul nécessaire à de tels modèles.

Par cette étude, nous souhaitons comparer les deux approches variationnelles entre elles, mais également comparer les algorithmes déterministes proposés avec deux algorithmes stochastiques de classification connus pour leur robustesse au bruit.

4.1 Introduction

Dans ce chapitre, nous comparons les résultats obtenus par les deux modèles variationnels de classification proposés, et nous les confrontons à deux modèles stochastiques présentés dans le chapitre 1. Nous comparons les deux modèles proposés entre eux, mais également avec deux approches stochastiques de complexité différente : le modèle de Potts, qui est relativement simple et que nous optimisons à l'aide d'un recuit simulé dans le cadre de l'estimateur du MAP ; le chien-modèle, qui est relativement complexe et qui est optimisé dans le cadre de l'estimateur du MPM.

Les deux modèles que nous avons développés reposent sur des approches différentes. Le premier aborde la classification comme un problème de restauration avec contrainte de classification. La solution est obtenue par minimisation d'une suite de fonctionnelles. Le second modèle proposé est la traduction d'un problème de partition à l'aide de contours et régions actifs dans un formalisme par ensembles de niveaux. La nature de ces deux modèles variationnels diffère fondamentalement, même s'ils aboutissent tous deux à des solutions de partition de l'image possédant des propriétés de régularité de même type.

Nous allons résumer les propriétés des fonctionnelles limites dans les deux cas. Rappelons l'expression des deux fonctionnelles (2.11) et (3.35) proposées :

$$J_\varepsilon(f) = \int_{\Omega} \left(f(x) - I(x) \right)^2 dx + \varepsilon \lambda^2 \int_{\Omega} \varphi(|\nabla f(x)|) dx + \frac{\eta^2}{\varepsilon} \int_{\Omega} W(f(x); \vec{\mu}, \vec{\sigma}) dx$$

et

$$F_\alpha(\Phi_1, \dots, \Phi_K) = \sum_{i=1}^K e_i \int_{\Omega} H_\alpha(\Phi_i) \frac{(I(x) - \mu_i)^2}{\sigma_i^2} dx + \sum_{i=1}^K \gamma_i \int_{\Omega} g(|\nabla I|) \delta_\alpha(\Phi_i) |\nabla \Phi_i| dx \\ + \frac{\lambda}{2} \int_{\Omega} \left(\sum_{i=1}^K H_\alpha(\Phi_i) - 1 \right)^2 dx.$$

La fonctionnelle J dépend du paramètre de Γ -convergence ε , alors que F comporte un paramètre α d'approximation des distributions de Dirac et de Heaviside. Lorsque ces deux paramètres tendent vers 0, nous obtenons deux fonctionnelles limites : J_0 et F_0 .

Supposons que $\varphi(t) = t^2$. D'après la théorie de Van der Waals-Cahn-Hilliard, nous avons vu que :

$$J_\varepsilon(f) \xrightarrow{\varepsilon \rightarrow 0^+} J_0(f) = \int_{\Omega} \left(f(x) - I(x) \right)^2 dx + \sum_{i,j=1}^K d(\mu_i, \mu_j) \mathcal{H}^1(\Gamma_{ij}) \begin{cases} f \in BV(\Omega), \\ f(x) \in [\mu_1, \dots, \mu_K] \text{ p.p} \end{cases} \quad (4.1)$$

Concernant la convergence de F_α , lorsque les approximations δ_α et H_α convergent vers les distributions de Dirac et de Heaviside, nous avons :

$$\sum_{i=1}^K e_i \int_{\Omega} H_\alpha(\Phi_i) \frac{(I(x) - \mu_i)^2}{\sigma_i^2} dx \xrightarrow{\alpha \rightarrow 0^+} \sum_{i=1}^K e_i \int_{\Omega_i} \frac{(I(x) - \mu_i)^2}{\sigma_i^2} dx \quad (4.2a)$$

$$\sum_{i=1}^K \gamma_i \int_{\Omega} g(|\nabla I|) \delta_\alpha(\Phi_i) |\nabla \Phi_i| dx \xrightarrow{\alpha \rightarrow 0^+} \sum_{i=1}^K \gamma_i \int_{\Gamma_i} g(|\nabla I|) d\Gamma_i \quad (4.2b)$$

$$\frac{\lambda}{2} \int_{\Omega} \left(\sum_{i=1}^K H_\alpha(\Phi_i) - 1 \right)^2 dx \xrightarrow{\alpha \rightarrow 0^+} \frac{\lambda}{2} \int_{\Omega} \left(\sum_{i=1}^K H(\Phi_i) - 1 \right)^2 dx \quad (4.2c)$$

donc :

$$F_\alpha(\Phi_1, \dots, \Phi_K) \xrightarrow{\alpha \rightarrow 0^+} F_0(\Phi_1, \dots, \Phi_K) = \sum_{i=1}^K e_i \int_{\Omega_i} \frac{(I(x) - \mu_i)^2}{\sigma_i^2} dx \\ + \sum_{i=1}^K \gamma_i \int_{\Gamma_i} g(|\nabla I|) d\Gamma_i \\ + \frac{\lambda}{2} \int_{\Omega} \left(\sum_{i=1}^K H(\Phi_i) - 1 \right)^2 dx \quad (4.3)$$

A travers l'expression des fonctionnelles limites, nous constatons que les solutions de minimisation de J_0 et F_0 seront des images étiquetées constituées de K ensembles de régions homogènes et repérées par des étiquettes, et dont les frontières sont de longueurs minimales¹. Les deux approches sont différentes mais on remarque que les contraintes que nous

¹la longueur des contours de la solution de minimisation de J_0 est pondérée par un terme tenant compte d'une certaine distance entre les puits

définissons pour chaque modèle fournissent des solutions de classification possédant le même type de propriétés. En pratique, nous effectuons une suite de minimisations pour J_ε . En revanche, comme nous l'avons vu au chapitre 3, le paramètre α est fixé à une petite valeur que nous ne faisons pas décroître. En effet, le nombre de pixels dans la bande définie par $B_\alpha^i = \{x \in \Omega / |\Phi_i(x)| \leq \alpha\}$ dans laquelle est mise à jour l'EDP (3.36) dépend de la valeur de α , compte tenu de la discrétisation adoptée et pour des fonctions de distance Φ_i (le pas de discrétisation entre deux pixels vaut 1).

4.2 Modèles de comparaison

Nous allons comparer entre eux les deux modèles variationnels que nous avons développés, et nous effectuons également des comparaisons avec deux approches stochastiques. De nombreux modèles stochastiques ont été proposés pour la classification et la segmentation d'images dans la littérature, comme cela a été mentionné au cours du premier chapitre. Les modèles probabilistes optimisés par des algorithmes stochastiques sont connus pour leur robustesse au bruit.

Nous rappelons que, dans ce cadre, nous cherchons un étiquetage probabiliste de l'image à travers la variable L .

Le maximum de vraisemblance (MV) consiste à minimiser l'énergie E de la forme :

$$\mathbf{MV} : \quad E(L) = \sum_{s=1}^N \left(\ln \sqrt{2\pi} \sigma_{L_s} + \frac{(O_s - \mu_{L_s})^2}{2\sigma_{L_s}^2} \right).$$

C'est-à-dire que nous n'effectuons pas de régularisation, nous calculons l'étiquette la plus proche en chaque pixel, au sens de la distance définie ci-dessus. Nous montrons les résultats obtenus par MV pour justifier l'introduction d'une régularisation dans les modèles de classification proposés.

Le modèle avec régularisation de type Potts présenté au paragraphe 1.3.2 est un des plus connus et des plus utilisés du fait de sa simplicité. Dans le cadre de l'estimateur du MAP, nous minimisons l'énergie suivante :

$$\mathbf{POTTS} : \quad E(L) = \sum_{s=1}^N \left(\ln \sqrt{2\pi} \sigma_{L_s} + \frac{(O_s - \mu_{L_s})^2}{2\sigma_{L_s}^2} \right) + \beta \sum_{\{s,r\} \in Q} \gamma(L_s, L_r).$$

Le voisinage utilisé est 4-connexe. Nous optimisons ce modèle stochastique à l'aide d'un recuit simulé avec une dynamique de Metropolis que nous avons exposé dans le tableau

1.5 au paragraphe 1.3.2. Cet algorithme, indépendant du choix des conditions initiales, est censé fournir le minimum global des énergies E définies plus haut, au contraire des algorithmes déterministes. Pour tous les tests qui suivent, nous avons initialisé l'algorithme de Metropolis par un étiquetage aléatoire de l'image. La décroissance de la température entre les itérations k et $k + 1$ est linéaire et faible : $T^{k+1} = \rho T^k$. Selon les applications, ρ a été choisi entre 0.95 et 0.99, et la température initiale T^0 entre 10 et 1.

Parmi les modèles stochastiques plus complexes, nous avons introduit le chien-modèle au paragraphe 1.3.2. Ce modèle permet de prendre en compte des contraintes sur des structures comme les lignes ou les contours grâce à des considérations sur les configurations que l'on peut rencontrer pour des cliques de taille 3×3 .

Nous optimisons la loi a posteriori suivante :

$$P(L/O) \propto P(O/L)P(L),$$

avec :

$$P(O/L) = \prod_{s \in S} P(O_s/L_s) = \prod_{s \in S} \frac{1}{\sqrt{2\pi}\sigma_{L_s}} \exp\left[-\frac{(O_s - \mu_{L_s})^2}{2\sigma_{L_s}^2}\right],$$

et l'a priori :

$$\textbf{CHIEN : } \quad P(L) = \frac{1}{Z(e, l, n)} \exp[-eN_0(L) - lN_1(L) - nN_2(L)].$$

Nous optimisons la loi a posteriori à l'aide d'un estimateur du MPM dont la fonction de coût correspondante est définie dans (1.14) et dont l'algorithme est présenté dans le tableau 1.14. Nous approchons la loi a posteriori à l'aide d'un échantillonneur de Gibbs [Besag, 1974, Geman et Geman, 1984], et nous gardons en chaque site l'étiquette la plus représentée parmi les derniers échantillons générés. Il est nécessaire d'effectuer un grand nombre de tirages afin d'obtenir des échantillons représentatifs de cette loi. Il en résulte un temps de calcul assez élevé pour cet estimateur.

Nous comparons les résultats obtenus par les modèles de classification que nous proposons au modèle de Potts, optimisé dans le cadre de l'estimateur du MAP, ainsi qu'au chien-modèle, optimisé par un estimateur de type MPM. Notons que nous pouvons trouver dans [Descombes *et al.*, 1997b, Descombes *et al.*, 1997a] une étude approfondie des modèles de Potts et du chien-modèle.

4.3 Comparaisons expérimentales

Par la suite, nous désignerons les différents résultats de classification par :

- **POTTS**, pour le modèle stochastique avec régularisation de Potts (cf. chapitre 1, paragraphe 1.3.2)
- **CHIEN**, pour le modèle stochastique avec un modèle a priori de type chien-modèle (cf. chapitre 1, paragraphe 1.3.2)
- **M1**, pour le premier modèle que nous avons proposé : classification avec restauration (cf. algorithme dans le tableau 2.2)
- **M2**, pour le second modèle que nous avons proposé : classification par modèle actif (cf. algorithme dans le tableau 3.2)

4.3.1 Présentation des résultats

Nous comparons les quatre modèles dont les abréviations sont définies ci-dessus sur deux images synthétiques ainsi que sur une image réelle SPOT. Soulignons le fait que seuls les résultats obtenus sur les données synthétiques sont étudiés en termes de nombre n de pixels mal classés (ou "erreurs") et de taux de réussite p , avec :

$$p = 100.(1 - \frac{n}{N}), \quad (4.5)$$

où N est le nombre de pixels de l'image considérée. En effet, ne disposant pas d'une vérité terrain pour les images réelles monospectrales à notre disposition, nous nous contenterons de montrer, à titre purement indicatif, les différentes classifications obtenues par les quatre modèles étudiés. Le chapitre 6 traite une application sur des données SPOT multispectrales pour lesquelles nous disposons d'une vérité terrain.

Image "gdr"

Cette image (voir figure 4.1), déjà traitée au sein des chapitres précédents, provient de la base de données du GDR-PRC ISIS. Nous avons ajouté un bruit gaussien de moyenne nulle donnant lieu à une version bruitée, de RSB égal à 10dB.

Nous avons mené des comparaisons entre les différents modèles de classification en définissant d'abord 3 classes, puis 8 classes. Les paramètres des classes sont donnés dans le tableau 2.3. Dans le cas de 8 classes, nous sommes en présence d'un mélange de classes dû à la proximité des moyennes ainsi qu'à la valeur des variances (cf. paramètres μ_i et σ_i dans le tableau 4.1). Les résultats de classification lorsque aucune erreur n'est commise sont présentés sur

la figure 4.1, et servent de vérité terrain (notée VT) pour juger de la qualité des résultats obtenus par les quatre modèles. Dans le cas de 8 classes, 6 d'entre elles sont présentes dans la zone en dégradé de niveaux de gris de l'image. Dans cette zone, les frontières verticales des 6 classes sont définies par les paramètres des classes, sans être liées à la présence d'un gradient élevé des niveaux de gris de l'image. En effet, dans le dégradé, les niveaux de gris sont décrémentés de 1 lorsqu'on se déplace horizontalement de gauche à droite (excepté au niveau des bords latéraux de la zone de dégradé). Dans ce sens, l'intensité lumineuse n'est pas stationnaire dans les classes.

Nous exposons sur la figure 4.2 les résultats obtenus par Maximum de Vraisemblance (MV) pour mettre en évidence l'importance de l'introduction d'une régularisation (modèle a priori) dans les modèles de classification utilisés.

La figure 4.3 présente les résultats obtenus par les différents modèles dans les cas de 3 classes et de 8 classes. Nous montrons sur la figure 4.4 le nombre de pixels mal classés et leur localisation ainsi que le pourcentage de réussite selon les différents modèles. Pour chacune des 8 classes, le nombre d'erreurs commises par les quatre modèles est exposé dans le tableau 4.1. Nous examinons plus précisément les différences entre les deux modèles de classification M1 et M2 que nous avons développés, en localisant sur la figure 4.5 les pixels différemment étiquetés par ces deux modèles.

Le temps de calcul ainsi que le nombre d'itérations nécessités par les quatre modèles sont exposés dans le tableau 4.2. Concernant le chien-modèle, nous indiquons entre parenthèses le nombre d'échantillons retenus pour estimer l'étiquette la plus représentée en chaque pixel (cf. algorithme du MPM dans le tableau 1.7).

Image "échiquier"

Nous avons ensuite étudié l'influence du niveau de bruit des données observées sur les résultats de classification. Nous avons pour cela généré trois versions bruitées (bruit gaussien et additif de moyenne nulle) dont les RSB sont de 10dB, 5dB et 0dB. Les données originales, ainsi que les trois versions bruitées, sont présentées sur la figure 4.6. Les moyennes des 4 classes de l'image sont identiques quel que soit le niveau de bruit (cf. valeurs de μ_i pour l'image "échiquier" dans le tableau 2.3), en revanche la valeur des écarts-types est celle du bruit introduit : pour un RSB de 10dB, $\sigma_i = 12.85$ ($i = 1, 2, 3, 4$) ; pour un RSB de 5dB, $\sigma_i = 22.5$ ($i = 1, 2, 3, 4$) ; pour un RSB de 0dB, $\sigma_i = 41.25$ ($i = 1, 2, 3, 4$). Pour les versions bruitées à 5dB et à 0dB, nous sommes en présence d'un mélange de classes dû à l'importance de la valeur de l'écart-type des classes.

Nous constatons d'après la figure 4.7 que le MV conduit à un nombre d'erreurs d'autant plus important que le niveau de bruit est élevé, d'où la nécessité d'injecter de l'information

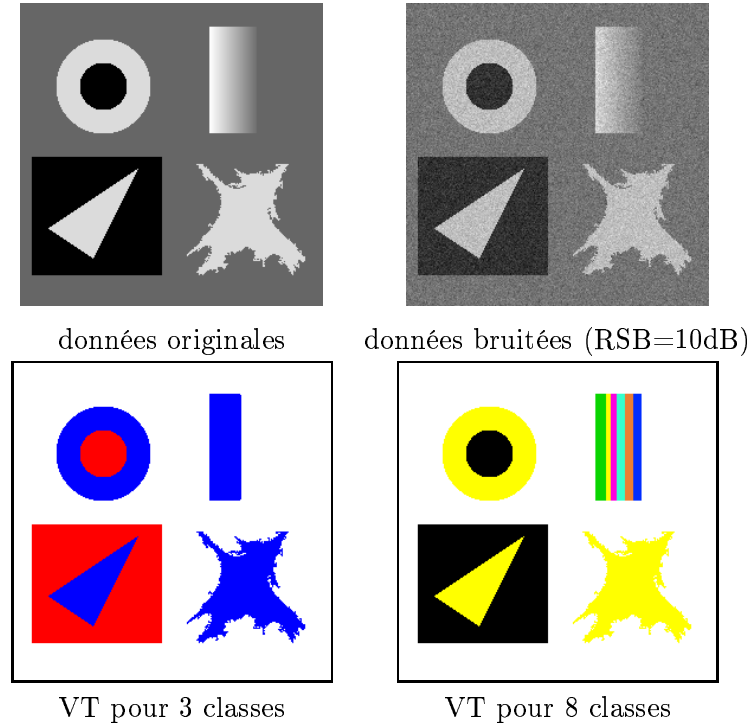


FIG. 4.1 – Image "gdr" bruitée et vérité-terrain (VT) pour 3 et 8 classes, image 256x256 pixels.

a priori sur la solution cherchée.

Nous présentons sur la figure 4.8 les résultats ainsi que les erreurs obtenus par les quatre modèles de classification en fonction du niveau de bruit sur les données.

4.3.2 Commentaires

Les expériences conduites permettent de mettre en évidence le comportement des différents modèles étudiés dans le cas où les frontières des classes ne sont pas liées à la présence de forts gradients des niveaux de gris, et lorsqu'il existe un mélange de classes. Le potentiel W du modèle M1 offre vraisemblablement une meilleure discrimination des classes que ne le permettent les termes d'attache aux données des trois autres modèles. De plus, l'algorithme d'optimisation utilisé pour M1 ne dépend pas du nombre de classes K (la seule conséquence porte sur le nombre de puits de W), contrairement à ceux utilisés pour optimiser M2 (système de K EDP à résoudre) et pour le chien-modèle (à cause de l'échantillonneur de Gibbs : calcul des K distributions conditionnelles locales). Cependant, lorsque le niveau de bruit est élevé (cf. figure 4.8), nous constatons que le modèle de contour utilisé pour M1 possède

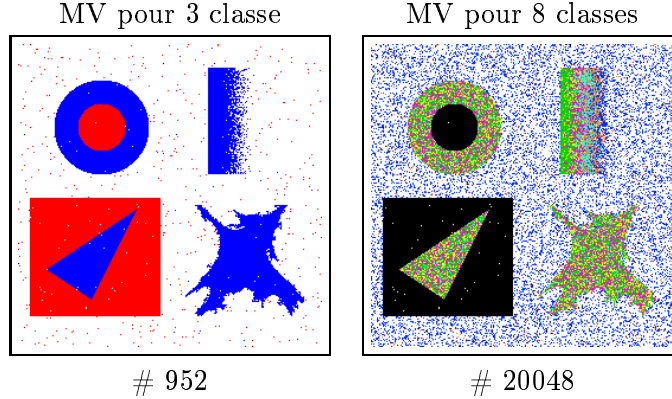










FIG. 4.2 – Maximum de vraisemblance : résultats de classification pour 3 et 8 classes ainsi que nombre de pixels mal classés (#.).

certaines lacunes. En effet, nous pouvons remarquer sur les résultats de l'image "échiquier" que les frontières des classes commencent à être lissées (malgré un lissage préservant les contours), alors qu'il subsiste de petites régions isolées liées à la présence de forts gradients et assimilés à des contours. Pour tous les tests effectués, le modèle M2 est initialisé de façon automatique par des petites bulles estimées dans des fenêtres 5×5 (cf. figure 3.7). Cette initialisation a un effet moyennant permettant à ce modèle d'être plus robuste à un niveau de bruit élevé que ne l'est M1, mais les résultats perdent en précision. On constate clairement ce phénomène sur les différents niveaux de bruit de l'image "échiquier" : la classification obtenue à 0dB ne comporte que très peu de pixels isolés. Nous pouvons remarquer qu'à 5dB, les régions dynamiques ne vont pas pénétrer dans les recoins de l'image possédant une forte courbure comme dans les carrés du damier intersectés par le disque. Le modèle de Potts, comme le chien-modèle, ne donnent pas de bons résultats dans le cas d'un mélange de classes comme celui rencontré pour l'image "gdr". Les paramètres des classes sont pris en compte au niveau du terme d'attache aux données, lequel est identique pour ces deux modèles. On constate dans le tableau 4.1 que le nombre d'erreurs est maximum, et ceci quels que soient les modèles, pour la classe 7 (contenant environ un pixel sur six de l'image) qui est très proche des moyennes des classes des classes 6 et 8. Les quatre modèles commettent le maximum d'erreurs pour la classes 6, 7, et 8. Le modèle M1 est celui qui commet le moins d'erreurs, les deux modèles stochastiques en commettant le plus. La classe 1 dont la moyenne est très éloignée des autres, ne pose pratiquement aucun problème quelque soient les modèles de classification. Par ailleurs, le chien-modèle commet beaucoup plus d'erreurs que les autres modèles pour ce qui est de la classe 2, cela peut s'expliquer par le fait que ces pixels mal classés, provenant principalement de la classe 1, sont liés à des pixels bruités et

CLASSES

1  $\mu_1 = 100$	2  $\mu_2 = 128$	3  $\mu_3 = 135$	4  $\mu_4 = 142$	5  $\mu_5 = 149$	6  $\mu_6 = 156$	7  $\mu_7 = 160$	8  $\mu_8 = 163$
--	--	--	--	---	--	--	--

									total
POTTS	1	75	130	88	141	110	3831	4	4380
CHIEN	0	424	148	69	181	151	4650	26	5649
M1	0	62	40	42	73	71	320	66	674
M2	1	75	89	65	106	84	1661	32	2113
nb pixels pour VT	10294	41641	630	630	630	450	10451	810	65536

TAB. 4.1 – Image "gdr" 8 classes : Pour chaque classe : nombre de pixels mal classés selon les modèles. La dernière ligne comptabilise le nombre de pixels d'échantillons par classe pour la vérité terrain (VT) de la figure 4.1.

entre lesquels le modèle a priori a favorisé la présence d'une ligne plutôt que celle d'une zone homogène. En revanche, nous constatons que sur les images "échiquier" à 5dB et à 0dB, les deux modèles stochastiques offrent de bons résultats, comparés aux modèles variationnels M1 et M2. Ceci s'explique par le caractère stochastique des algorithmes d'optimisation qui permettent de sortir de certains minima locaux des énergies associées. Pour les données bruitées à 0db, le chien-modèle, dont la complexité est bien supérieure à celle du modèle de Potts, et dont l'optimisation est menée à travers un estimateur MPM, offre les meilleurs résultats en termes de nombre d'erreurs. Cependant, cela se paie par un temps de calcul relativement élevé par rapport aux autres modèles (voir le tableau 4.2 pour les données "gdr").

4.4 Conclusion

Les expériences conduites dans le cas d'un mélange de classes et pour différents niveaux de bruits introduits sur des données synthétiques permettent de tirer les conclusions suivantes :

- Jusqu'à un certain niveau de bruit, le modèle variationnel M1, combinant restauration et classification, s'avère être performant à tous les niveaux (taux de réussite, gestion du mélange de classes, temps calculatoire), que ce soit dans le cas d'un mélange de classes ou

image "gdr" 3 CLASSES

modèle	temps	nbre itérations
M1	49 s.	16
M2	242 s	190
POTTS	385 s.	1000
CHIEN	1738 s.	800(400)

image "gdr" 8 CLASSES

modèle	temps	nbre itérations
M1	101 s.	35
M2	747 s.	280
POTTS	821 s.	2000
CHIEN	3960 s.	1200(600)

TAB. 4.2 – Temps d'exécution et nombre d'itérations des différents modèles de classification pour l'image "gdr" avec 3 et 8 classes (temps cpu en secondes, processeur 300 Mhz). Ces 3 images ont une taille de 256x256 pixels.

dans le cas contraire.

- Pour des données fortement bruitées, le modèle de contour considéré dans M1 étant trop simple, ce modèle de classification conduit à un grand nombre d'erreurs. En revanche, le modèle variationnel dynamique M2 se comporte mieux en présence d'un fort bruit grâce à un modèle de contour plus élaboré et à une initialisation automatique ayant un effet moyennneur.

- Les modèles stochastiques de Potts et du chien commettent beaucoup d'erreurs dans le cas d'un mélange de classes du type de celui rencontré pour l'image "gdr", malgré le caractère stochastique des algorithmes de minimisation associés qui leur permettent d'éviter de converger vers une solution locale. En revanche, lorsque les classes ne sont pas mélangées et en présence d'un fort bruit sur les données, ces deux modèles, donnent de meilleurs résultats que ne le permettent les modèles variationnels proposés, au détriment d'un coût calculatoire plus élevé. Le chien-modèle obtient de meilleurs résultats que le modèle de Potts, grâce à un a priori plus complexe sur la solution cherchée, et à un estimateur MPM plus performant que le MAP dans ce cas, mais bien plus long à converger.

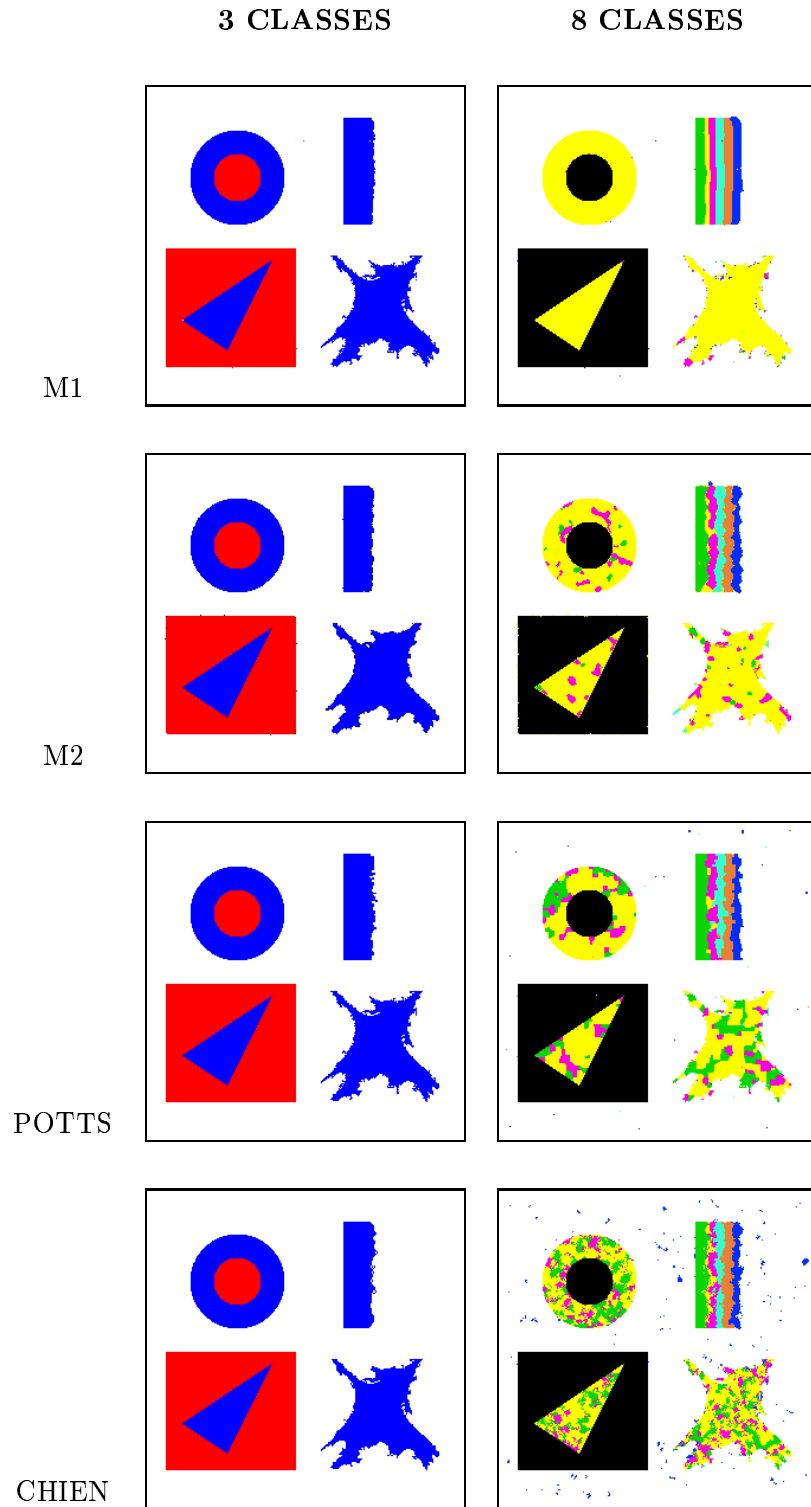


FIG. 4.3 – Image "gdr" bruitée : résultats de classification pour 3 classes et 8 classes selon les différents modèles.

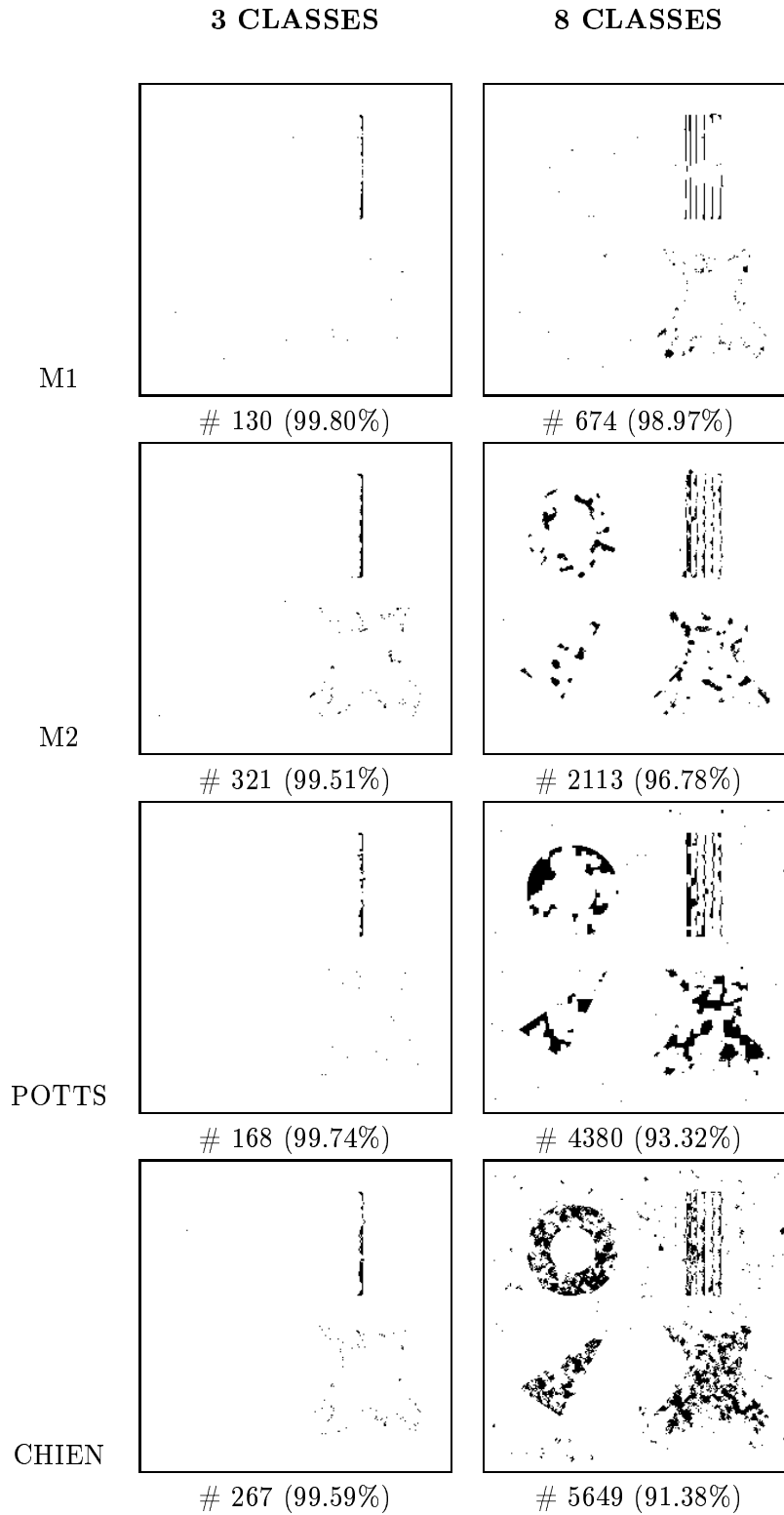


FIG. 4.4 – Nombre (#.) et localisation des pixels mal classés par les 4 modèles sur l'image "gdr" pour 3 et 8 classes. Le pourcentage de réussite est indiqué entre parenthèses.

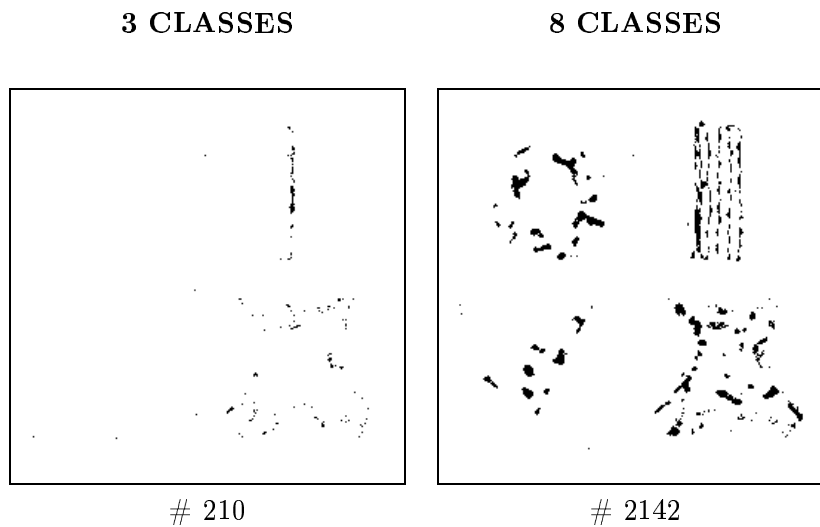


FIG. 4.5 – Nombre (#.) et localisation des pixels différemment classés par les modèles M1 et M2 sur l'image "gdr" pour 3 et 8 classes.

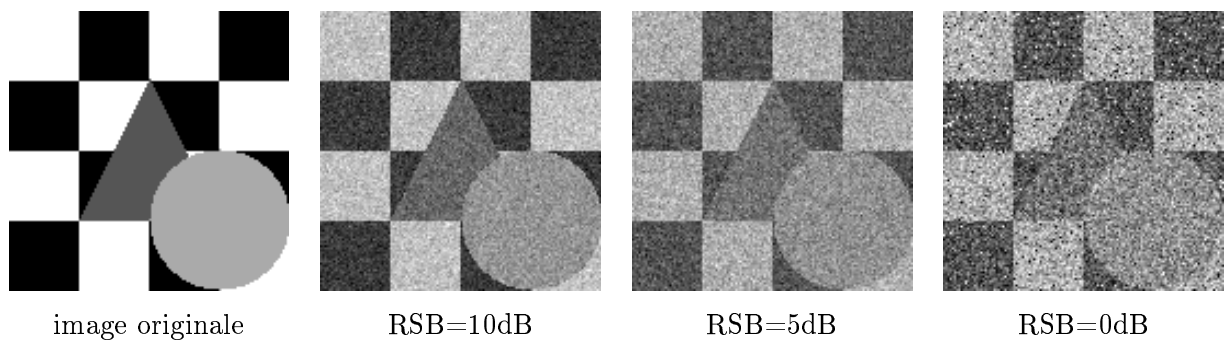


FIG. 4.6 – Image "échiquier" (128×128 pixels) originale, et les trois versions bruitées.

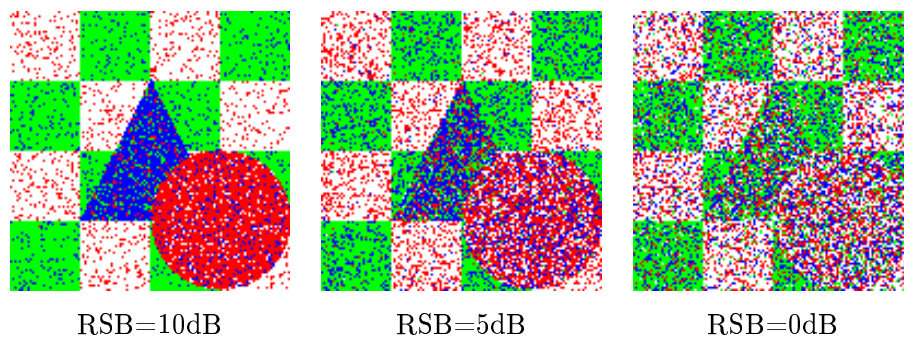


FIG. 4.7 – Classification de l'image "échiquier" par MV (4 classes) pour différents niveaux de bruit.

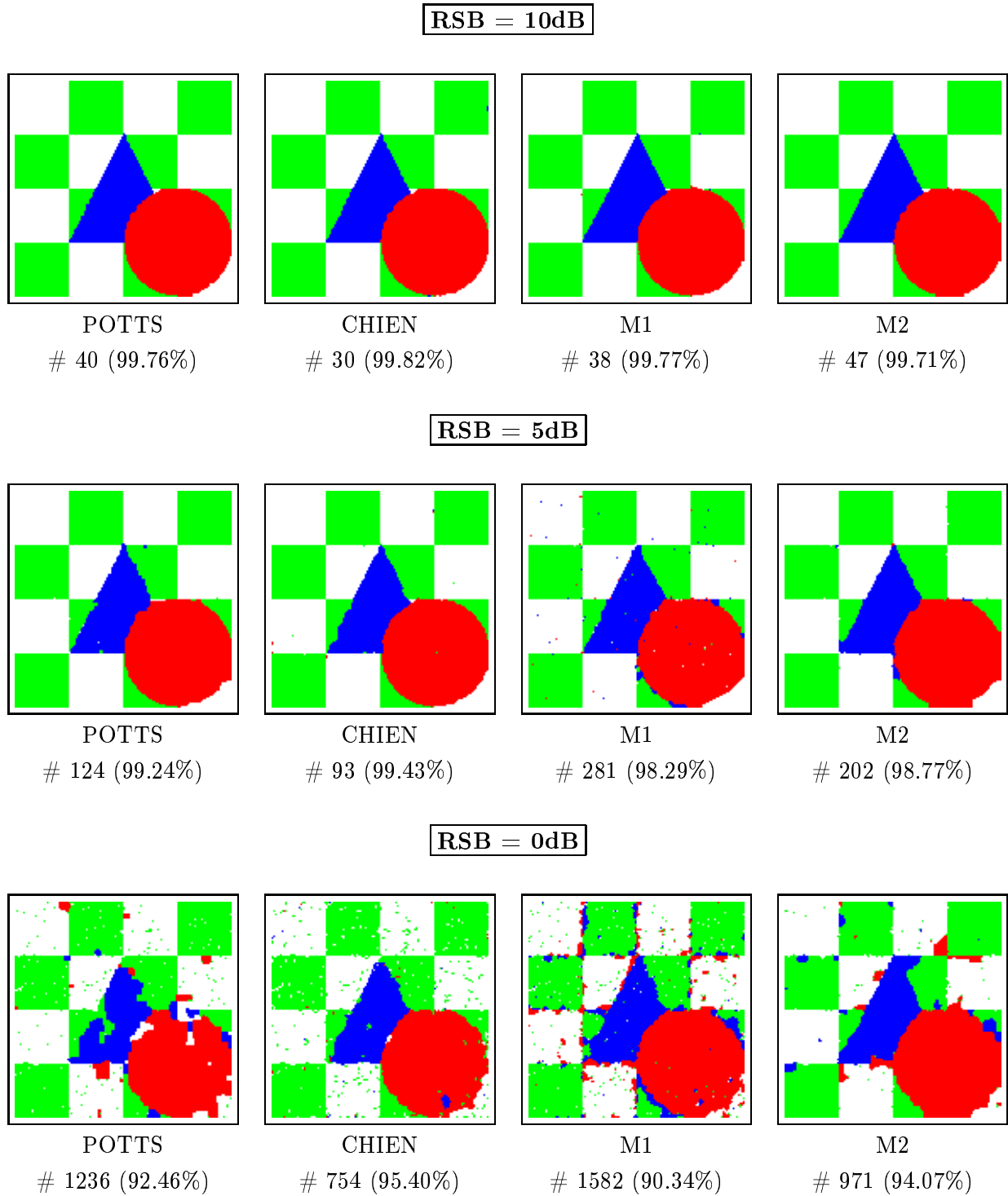


FIG. 4.8 – Classification de l'image "échiquier" (4 classes) pour différents niveaux de bruit et selon les modèles. Nous indiquons le nombre de pixels mal classés ainsi que le pourcentage de réussite.

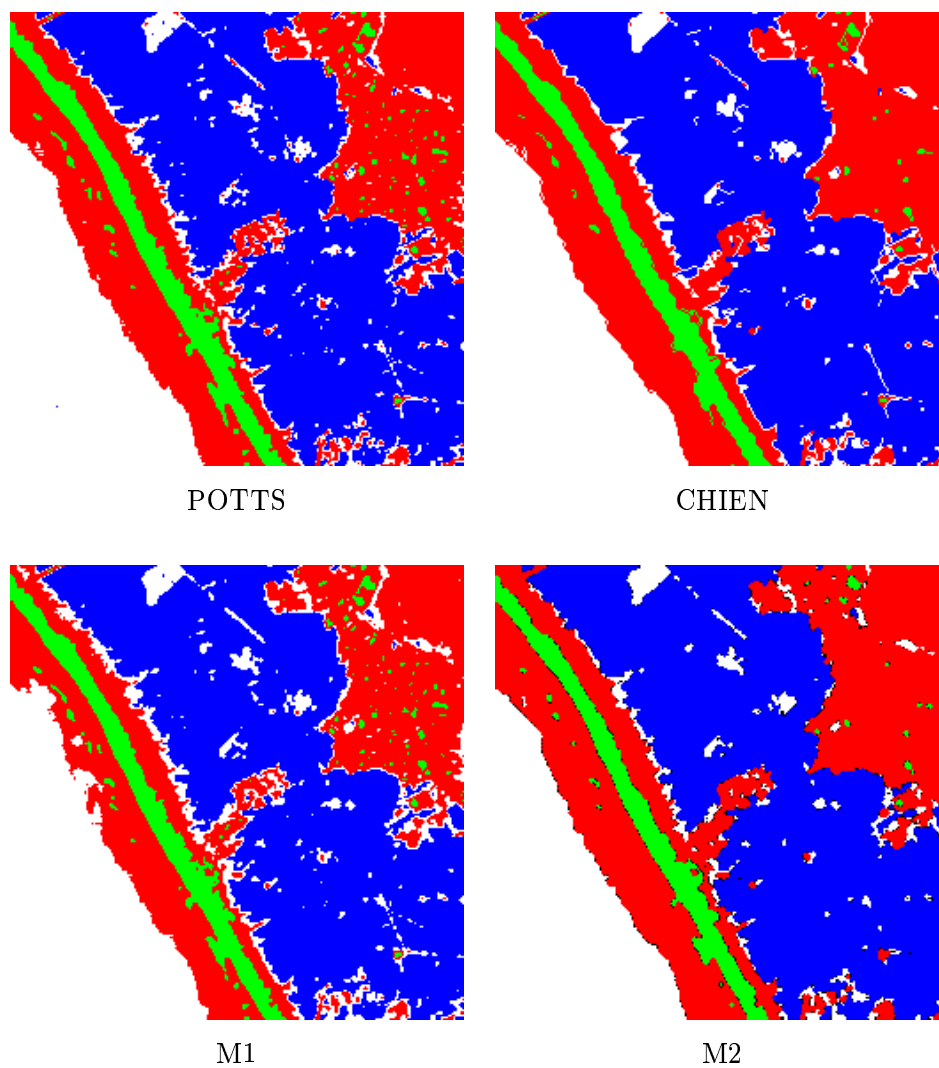
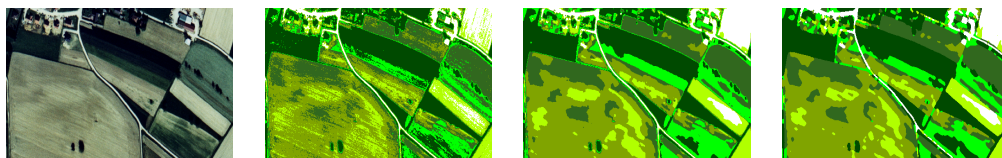


FIG. 4.9 – Résultats de classification pour l'image SPOT présentée sur la figure 2.17 du chapitre 2.

Chapitre 5

Le cas multispectral



Où nous étendons les deux modèles de classification présentés aux chapitres 2 et 3 aux cas de données multispectrales. Nous traitons les différentes images renvoyées par le satellite afin de combiner l'information contenue dans les différentes bandes spectrales. Le couplage entre les bandes intervient à des niveaux différents pour les deux modèles. Concernant le modèle reposant sur la minimisation d'une suite de fonctionnelles, le couplage entre les bandes s'effectue au niveau du calcul du gradient sur les données, ainsi qu'à travers du potentiel de contrainte de niveaux. Pour le modèle de régions/contours actifs, la combinaison d'information fournie par les différentes bandes intervient au niveau du terme d'attache aux données.

5.1 Introduction

Nous introduisons dans un premier temps les particularités du cas multispectral, en insistant sur les points concernés par l'extension des deux modèles présentés au sein des chapitres 2 et 3 au cas de données multiples. L'extension au cas multispectral intervient à des niveaux différents pour ces deux modèles. Nous commençons par exposer les différentes notations utilisées dans le cas multispectral, ainsi que par une présentation du modèle d'estimation du nombre et des paramètres des classes que nous utilisons au préalable, avant d'aborder l'extension de chacun des deux modèles au cas multispectral.

5.1.1 Le cas multispectral

Ce chapitre concerne la classification d'images multispectrales à l'aide des modèles proposés dans les chapitres précédents. Il s'agit de combiner l'information retournée dans les différentes bandes spectrales par les satellites. Nous ne traiterons pas de façon approfondie de la redondance d'information entre les bandes (voir par exemple [Volden *et al.*, 1996a, Volden *et al.*, 1996b, Winter, 1997]) pour laquelle il s'agit de prendre en compte la physique des capteurs, même si nous savons par exemple qu'il existe une forte redondance d'information entre les bandes pour les images multispectrales fournies par les satellites SPOT. La corrélation entre les bandes est prise en compte par l'intermédiaire de matrices de covariances.

Dans ce chapitre, nous supposerons que les données observées sont représentées par une fonction constituée de P composantes : les P signaux renvoyés par les différents capteurs des satellites. Nous utiliserons les notations suivantes :

- P est le nombre de bandes considéré.
- Les observations sont représentées par la fonction $\vec{I} : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}^P$, soit $\vec{I} = [I^1, \dots, I^P]^T$. Nous supposons que $\vec{I} \in L^2(\Omega, \mathbb{R}^P)$.

- Nous supposons que le nombre de classes K est le même dans chaque bande, chaque bande représentant la même scène dans différents domaines spectraux.
- Les K classes sont caractérisées par les paramètres gaussiens suivants :

$$\vec{\mu}_i = \{\mu_i^p\}_{p=1\dots P} \quad \text{et} \quad \vec{\sigma}_i = \{\sigma_i^p\}_{p=1\dots P} \quad \text{pour} \quad i = 1, \dots, K.$$

- Pour $i = 1, \dots, K$, nous noterons Σ_i la matrice de variance-covariance définie sur $\mathbb{R}^P \times \mathbb{R}^P$ par :

$$\Sigma_i = \begin{bmatrix} \sigma_i^{11} & \sigma_i^{12} & \dots & \sigma_i^{1P} \\ \sigma_i^{21} & \sigma_i^{22} & \dots & \sigma_i^{2P} \\ \vdots & & \ddots & \vdots \\ \sigma_i^{P1} & \sigma_i^{P2} & \dots & \sigma_i^{PP} \end{bmatrix}.$$

où $\sigma_i^{mn} = \text{cov}(I^m(x), I^n(x))$ et avec $[\sigma_i^p]^2 = \sigma_i^{pp}$ la variance associée à la classe C_i pour la bande p .

Bien que nous n'étudions pas en détails la redondance d'information entre les P bandes, les covariances sont, en quelque sorte, une mesure de la corrélation existant entre les différentes bandes pour chacune des K classes. Une valeur élevée de σ_i^{mn} signifie que l'information sur la classe C_i dans la bande m est fortement corrélée à celle dans la bande n . Une matrice Σ_i diagonale, signifie que les bandes sont complètement décorrélées concernant la classe C_i .

Etant données les P bandes spectrales, il s'agit de trouver une partition des données, composée de K régions homogènes séparées par des interfaces de longueurs minimales. Nous attendons ainsi les mêmes propriétés sur les solutions de classification fournies par les deux modèles présentés dans les chapitres précédents, que ce soit dans les cas monospectral ou multispectral.

5.1.2 Estimation du nombre de classes et de leurs paramètres

L'étape d'estimation du nombre de classes et de leurs paramètres peut être effectuée au préalable à l'aide d'un algorithme de *Fuzzy Cmeans* [Bezdek, 1981, Krishnapuram et Keller, 1994, Rostaing *et al.*, 1999] dont le critère a été modifié par A. Lorette dans [Lorette, 1999, Lorette *et al.*, 2000] par l'ajout d'un terme d'entropie.

Cette méthode entre dans le cadre de la partition d'ensembles flous tels qu'ils ont été définis dans [Zadeh, 1965]. Dans ce contexte, il s'agit de découper l'image en un ensemble de Cl nuages de points (les ensembles *flous*, ou encore les ensembles de régions associés aux Cl classes) et d'attribuer à chaque site $s \in S$ de l'image, Cl nombres compris entre 0 et 1 représentant le degré d'appartenance à chacun des nuages. La partition de l'image est alors

caractérisée par une matrice $A = [a_s^i]$ où $a_s^i \in [0; 1]$ représente le degré d'appartenance du site s au nuage $i \in [1, 2, \dots, Cl]$.

Un terme d'entropie est utilisé pour estimer le nombre de classes (voir [Lorette *et al.*, 2000, Lorette, 1999] pour plus de détails). Le critère obtenu s'écrit :

$$J = \sum_{i=1}^{Cl} \sum_{s \in S} (a_s^i)^m d^2(I_s, c_i) - \alpha \sum_{i=1}^{Cl} p_i \log(p_i), \quad (5.1)$$

sous la contrainte

$$\sum_{i=1}^{Cl} a_s^i = 1 \quad \forall s \in S.$$

Le paramètre $m \in [0; \infty]$ caractérise le degré de flou. En pratique, il est fréquent de fixer $m = 2$ pour bon nombre d'applications (pour $m = 0$, on retrouve la méthode classique du *K means* [MacQueen, 1967] plus un terme d'entropie). Les valeurs $c_i = (\vec{\mu}_i, \vec{\sigma}_i)$ représentent les vecteurs caractéristiques des centres de gravité des nuages et $d^2(I_s, c_i)$ est la distance euclidienne entre l'observation I_s et c_i . La probabilité a priori p_i de la $i^{\text{ème}}$ classe est définie par

$$p_i = \frac{1}{N} \sum_{s \in S} a_s^i, \quad \text{avec } N = \text{card}(S).$$

Le premier terme de J caractérise l'homogénéité des nuages ; il est minimum lorsqu'un nuage ne caractérise qu'un seul niveau de gris. Le second terme, entropique, est lié au nombre de nuages, il est minimum lorsque tous les pixels sont dans le même nuage et que tous les autres sont vides.

En injectant la contrainte $\sum_{i=1}^{Cl} a_s^i = 1 \quad \forall s \in S$ sous forme de pénalité dans le critère, nous sommes conduits à la minimisation par rapport aux c_i et par rapport aux a_s^i du critère :

$$E = \sum_{i=1}^{Cl} \sum_{s \in S} (a_s^i)^m d^2(I_s, c_i) - \alpha \sum_{i=1}^{Cl} p_i \log(p_i) - \sum_{s \in S} \lambda_s \left[\sum_{i=1}^{Cl} a_s^i - 1 \right]. \quad (5.2)$$

Les paramètres λ_s représentent les multiplicateurs de Lagrange dans le cadre de l'optimisation sous contrainte.

Nous alternons les minimisations de E par rapport aux c_i (les a_s^i étant fixés) et par rapport aux a_s^i (avec les c_i fixés), en annulant les dérivées partielles associées.

Il est préférable de commencer l'algorithme avec une grande valeur de α afin de déterminer dès les premières itérations le nombre optimal de nuages de points (i.e. le nombre de classes K que nous allons considérer dans les modèles multispectraux détaillés dans les paragraphes suivants), puis de tenir compte peu à peu du terme de classification. Pour ce faire, α décroît

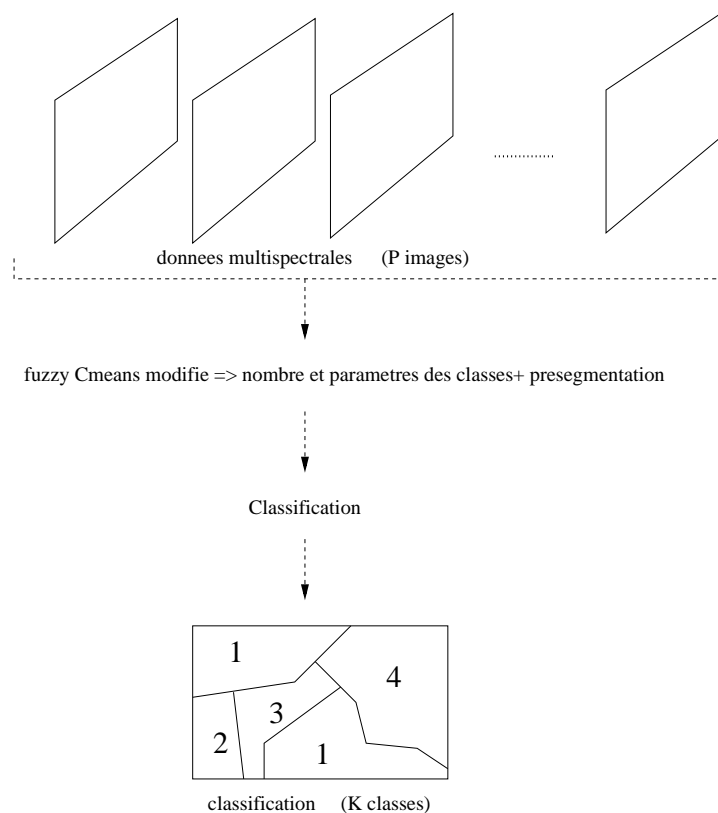


FIG. 5.1 – Principe de la classification automatique dans le cas multispectral.

de manière exponentielle [Frigui et Krishnapuram, 1996] à chaque itération de la manière suivante : $\alpha(n) = \alpha(0) \exp\left(-\frac{n}{\tau}\right)$, où n est le nombre d'itérations et τ est une constante de temps.

On évite ainsi que le terme de classification ne soit biaisé par le terme d'entropie puisque $\alpha(n)$ tend vers 0 quand n tend vers ∞ .

En outre, $\alpha(0)$ dépend fortement du nombre de pixels dans l'image. Afin de s'affranchir des données, nous posons $\alpha(0) = \alpha_0 N$ (où N est le nombre de pixels de l'image). Dans ce cas, le seul paramètre que nous avons à fixer est α_0 . Ce paramètre détermine la finesse de la classification que l'on souhaite obtenir. Pour une application donnée, nous fixons une fois pour toute α_0 et τ alors que le nombre de nuages de points initial Cl peut être différent suivant la scène traitée.

Après avoir mis à jour les centres de gravité et les degrés d'appartenance, nous estimons les probabilités a priori de chacun des nuages de points. Nous éliminons les nuages de points dont la probabilité est inférieure à un certain seuil ϵ . L'algorithme de *Fuzzy Cmeans* modifié est présenté dans le tableau 5.1.

L'algorithme de *Fuzzy Cmeans* modifié (issu des travaux de [Lorette, 1999]) permet donc

```

0 - Initialiser les centres de gravité des nuages de points,  $n = 0$ .
1 - Calculer la matrice  $A^{(n)}$ 
2 - Calculer les centres de gravité des nuages de points
3 - Calculer les probabilités a priori  $p_i$  de chaque nuage de points :
    si  $p_i < \epsilon$  le nuage de points  $i$  est supprimé
4 - Si  $\|A^{(n)} - A^{(n-1)}\| > \xi$  alors :
     $n = n + 1$ 
    mettre à jour  $\alpha$ 
    retourner à l'étape 1
    - Sinon aller à l'étape 5
5 - Assigner à chaque pixel l'étiquette du nuage de points auquel
    il a la plus grande probabilité d'appartenir.
```

TAB. 5.1 – Algorithme de *Fuzzy Cmeans* modifié.

de trouver le nombre de classes K , ainsi qu'une première classification sur laquelle nous pouvons estimer les paramètres gaussiens de chaque classe. Le nombre de classes, ainsi que leur paramètres sont ensuite donnés en entrée des modèles multispectraux que nous proposons pour obtenir une classification finale, comme l'illustre la figure 5.1. En pratique, nous utilisons l'algorithme de *Fuzzy Cmeans* modifié pour l'estimation des paramètres K , $\vec{\mu}_i$ et $\vec{\sigma}_i$, sans généralement utiliser le résultat de classification fourni par cet algorithme comme initialisation de nos modèles de classification multispectrale car nous initialisons le premier modèle proposé par les données observées et le second par une initialisation automatique dont le principe est présenté sur la figure 3.7.

5.2 Extension du premier modèle

5.2.1 Introduction

Nous proposons une extension du modèle de restauration avec contrainte de classification présenté au chapitre 2 pour le cas de données $\vec{I}: \mathbb{R}^2 \rightarrow \mathbb{R}^P$.

Dans le cadre des travaux portant sur la théorie de Van der Waals-Cahn-Hilliard, rappelons

le problème P_ε défini dans (2.7), dans le cas où $u : \mathbb{R}^2 \rightarrow \mathbb{R}$ et pour $W : \mathbb{R} \rightarrow \mathbb{R}^+$:

$$P_\varepsilon \begin{cases} \inf_{u \in H^1(\Omega; \mathbb{R})} E_\varepsilon(u), \\ E_\varepsilon(u) = \int_\Omega \left[\varepsilon |\nabla u(x)|^2 + \frac{1}{\varepsilon} W(u(x)) \right] dx, \\ \text{avec la contrainte suivante sur } u : \int_\Omega u(x) dx = m, \end{cases} \quad (5.3)$$

L'étude théorique de la Γ -convergence de E_ε lorsque $\varepsilon \rightarrow 0^+$ et de la caractérisation des suites de solutions u_ε a été étendue au cas où $u : \mathbb{R}^2 \rightarrow \mathbb{R}^P$ et pour $W : \mathbb{R}^P \rightarrow \mathbb{R}^+$ par [Baldo, 1990].

Dans le cas vectoriel, les conditions imposées sur le potentiel W sont de même nature que celles nécessaires dans le cas scalaire, à savoir :

- W s'annule sur les K minima $\vec{\mu}_i$ de \mathbb{R}^P ,
- W est localement quadratique autour de chaque minimum,
- W croît au moins linéairement en l'infini.

Les résultats montrés par S. Baldo dans le cas vectoriel sont du même type que dans le cas scalaire (cf. paragraphe 2.3 du chapitre 2), en particulier la solution limite $\lim_{\varepsilon \rightarrow 0^+} u_\varepsilon$ est géométriquement caractérisée par une partition en régions homogènes de niveaux μ_i^p pour la bande p , séparées par des interfaces de longueurs minimales. Les résultats présentés dans le paragraphe 2.3 du chapitre 3 sont ainsi valables dans le cas de variables vectorielles $u : \mathbb{R}^2 \rightarrow \mathbb{R}^P$ et $W^P : \mathbb{R}^P \rightarrow \mathbb{R}^+$.

Dans le cas vectoriel et pour K minima, l'expression de E_0 , la Γ -limite de E_ε définie dans (5.3) est exprimée dans [Baldo, 1990] par :

$$E_0(u) = \sum_{i,j=1}^K d(\vec{\mu}_i, \vec{\mu}_j) \mathcal{H}^1(\Gamma_{ij}) \text{ si } E_0(u) < +\infty, \ u \in BV(\Omega), \ W(u(x)) = 0 \text{ p.p.} \quad (5.4)$$

avec Γ_{ij} l'interface entre les régions Ω_i de la classe i et Ω_j . La distance d est définie par :

$$d(\vec{\mu}_i, \vec{\mu}_j) = \inf_g \left\{ \int_0^1 \sqrt{W(g(s))} |g'(s)| ds; \right. \\ \left. g \in C^1([0; 1], \mathbb{R}^P) \text{ et } g > 0, \ g(0) = \vec{\mu}_i, \ g(1) = \vec{\mu}_j \right\} \quad (5.5)$$

Ainsi, le problème limite de minimisation de E_ε est un problème de minimisation de la longueur pondérée de l'ensemble des interfaces entre les régions associées à chaque classe.

Si u_ε , solution de P_ε , converge dans $L^1(\Omega; \mathbb{R}^P)$ vers u_0 lorsque $\varepsilon \rightarrow 0^+$, alors :

$$u_0(x) = \sum_{i=1}^K \vec{\mu}_i \chi_{\Omega_i}(x) \quad (5.6)$$

avec χ_{Ω_i} la fonction caractéristique de l'ensemble Ω_i .

Par conséquent, d'après (5.6), la suite de solutions u_ε de la famille de problèmes P_ε définie par (5.6) est une partition régulière de Ω qui minimise la quantité $\sum_{i,j=1}^K d(\vec{\mu}_i, \vec{\mu}_j) \mathcal{H}^1(\Gamma_{ij})$. D'où les propriétés de régularité sur la solution du problème limite annoncées plus haut et démontrées formellement par S. Baldo.

Comme au paragraphe 2.3.2, nous allons faire un parallèle entre la théorie de Van der Waals-Cahn-Hilliard et le traitement d'images moyennant quelques extrapolations dues aux spécificités des images et du traitement de données réelles (comme l'utilisation d'un terme de perturbation sous-quadratique à la place de $|\nabla u(x)|^2$).

5.2.2 Modèle multispectral

Les principales difficultés liées à l'extension du modèle (2.11) au cas multispectral proviennent de l'expression du module du gradient de f dans \mathbb{R}^P , ainsi que de la construction du potentiel W sur \mathbb{R}^P . La famille de fonctionnelles que nous proposons de minimiser par rapport à $\vec{f}: \mathbb{R}^2 \rightarrow \mathbb{R}^P$ s'écrit :

$$J_\varepsilon(\vec{f}) = \underbrace{\int_{\Omega} |\vec{f}(x) - \vec{I}(x)|_{\mathbb{R}^P}^2 dx}_{\text{attache aux données}} + \varepsilon \lambda^2 \underbrace{\int_{\Omega} \varphi(|\nabla \vec{f}(x)|_{\mathbb{R}^P}) dx}_{\text{terme de restauration}} + \frac{\eta^2}{\varepsilon} \underbrace{\int_{\Omega} W(\vec{f}(x)) dx}_{\text{contrainte de niveaux}}, \quad (5.7)$$

et le problème associé consiste à trouver \tilde{f} vérifiant :

$$\tilde{f} = \lim_{\varepsilon \rightarrow 0^+} \left[\arg \min_{\vec{f}} J_\varepsilon(\vec{f}) \right]. \quad (5.8)$$

avec $|\cdot|_{\mathbb{R}^P}$ désignant une norme dans \mathbb{R}^P .

Le potentiel W prenant comme argument le vecteur \vec{f} , s'annule aux K points $\vec{\mu}_i$, minima du potentiel, et W possède une forme gaussienne (dans \mathbb{R}^P) autour de chaque minima, comme l'illustre la figure 5.2.

Comme pour le cas monospectral, nous utilisons une fonction de régularisation φ , non nécessairement quadratique, ce qui constitue un écart par rapport au cadre théorique défini dans [Baldo, 1990].

L'information des différentes bandes est combinée au niveau du gradient appliqué à \vec{f} , ainsi qu'à travers le potentiel de classification W .

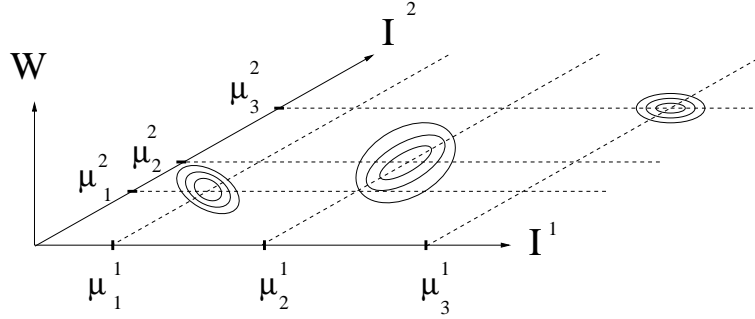


FIG. 5.2 – Projection du potentiel vectoriel W sur le plan (I^1, I^2) dans le cas de 2 bandes et pour 3 classes.

Calcul des normes dans \mathbb{R}^P

Les normes de \mathbb{R}^P définies dans (5.7) ne sont pas de même nature, qu'il s'agisse du terme d'attache aux données ou de l'argument de la fonction régularisante :

- $\left\| \vec{f}(x) - \vec{I}(x) \right\|_{\mathbb{R}^P}^2$: la norme portant sur le terme d'attache aux données est la norme $L^2(\mathbb{R}^P)$.
- $|\nabla \vec{f}(x)|_{\mathbb{R}^P}$: nous supposons pour le moment que nous utilisons la norme $L^2(\mathbb{R}^P)$.

Extension du potentiel W

Le potentiel W , reliant la solution \vec{f} aux classes et à leurs caractéristiques, doit tenir compte de la multi-dimensionalité des moyennes et variances dans le cas multispectral.

Nous allons prendre en compte les covariances des classes en plus des moyennes et variances caractéristiques, afin de combiner l'information sur les paramètres des classes à travers les différentes bandes. Ainsi, après avoir pré-estimé le nombre de classes K par un algorithme de *fuzzy Cmeans* modifié par exemple (cf. paragraphe 5.1.2), nous calculons leurs moyennes $\vec{\mu}_i \in \mathbb{R}^P$ et écart-types $\vec{\sigma}_i \in \mathbb{R}^P$ pour $i = 1, \dots, K$, ainsi que les covariances σ_i^{mn} pour m et $n = 1, \dots, P$.

Comme nous l'avons précisé précédemment, W est localement quadratique autour de chaque minimum $\vec{\mu}_i$, et s'annule en ces points. Nous définissons alors l'expression locale de W par :

$$W_{loc}(\vec{f}) = [\vec{f}(x) - \vec{\mu}_i]^T \Sigma_i^{-1} [\vec{f}(x) - \vec{\mu}_i] \quad (5.9)$$

avec Σ_i^{-1} désignant l'inverse de la matrice de variance-covariance (cf. paragraphe 5.1.1 pour les notations et définitions des covariances). Les covariances permettent de prendre en compte l'information sur une classe à travers les différentes bandes. Nous combinons ainsi l'information des différentes composantes de \vec{f} à travers les covariances.

Pour construire le potentiel W en fonction de chaque valeur de \vec{f} , il s'agit de déterminer le plus *proche* minimum $\vec{\mu}_i$ afin de déterminer l'expression locale (5.9). Pour cela, à chaque valeur de $\vec{f}(x)$, nous déterminons l'indice i tel que :

$$i = \arg \min_{k=1,\dots,K} [\vec{f}(x) - \vec{\mu}_k]^T \Sigma_k^{-1} [\vec{f}(x) - \vec{\mu}_k]. \quad (5.10)$$

La figure 5.3 illustre graphiquement, dans le cas d'une bande et pour 3 classes, la méthode utilisée d'après (5.10) permettant de sélectionner le plus *proche* minimum $\vec{\mu}_i$ pour chaque valeur de $\vec{f}(x)$, et d'en déduire l'expression locale de W (5.9). Sur cette figure, nous trouvons $i = 1$, et par conséquent, en ce point x nous avons :

$$W_{loc}(\vec{f}) = [\vec{f}(x) - \vec{\mu}_1]^T \Sigma_1^{-1} [\vec{f}(x) - \vec{\mu}_1].$$

Nous remarquons que W n'est pas dérivable aux points de jonctions des paraboles multidimensionnelles. Il faudrait pour cela effectuer par exemple des jonctions $C^1(\mathbb{R}^P)$, afin d'éviter d'éventuelles instabilités numériques pour des valeurs de $\vec{f}(x)$ se trouvant au niveau des points de jonction, ce que nous n'avons pas rencontré pour le moment au cours des applications traitées. La construction d'un potentiel W qui soit élément de $C^1(\mathbb{R}^P)$ fait partie des travaux à effectuer dans un proche avenir.

Fonctionnelle proposée

Compte tenu des calculs de normes multidimensionnelles adoptés et des expressions de W en fonction des valeurs de $\vec{f}(x)$, nous proposons de minimiser la fonctionnelle suivante par rapport à $\vec{f} \in \mathbb{R}^P$:

$$J_\varepsilon(\vec{f}) = \int_{\Omega} \left| \vec{f}(x) - \vec{I}(x) \right|_{\mathbb{R}^P}^2 dx + \varepsilon \lambda^2 \int_{\Omega} \varphi(|\nabla \vec{f}|_{\mathbb{R}^P}) dx + \frac{\eta^2}{\varepsilon} \int_{\Omega} W(\vec{f}(x)) dx, \quad (5.11)$$

avec W le potentiel de contrainte de niveaux dont l'expression locale est déterminée par (5.9).

Cette fonctionnelle combine l'information des différentes bandes au niveau du terme régularisant pondéré par ε ainsi qu'à travers le potentiel W . Nous utiliserons la géométrie riemannienne pour le calcul de $|\nabla \vec{f}|_{\mathbb{R}^P}$ uniquement au niveau des EDP d'Euler-Lagrange.

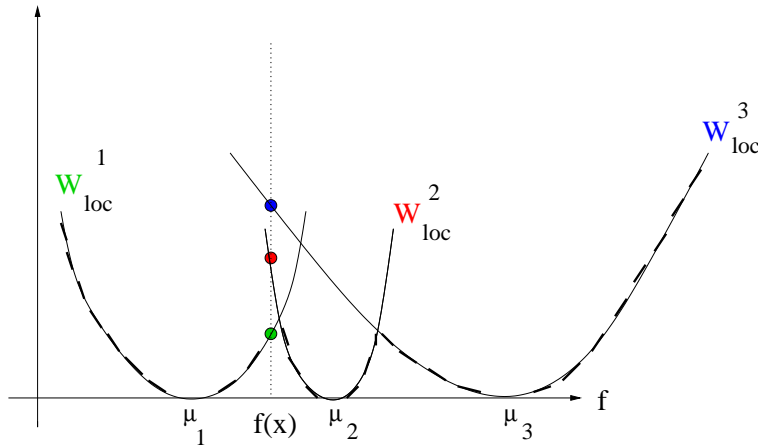


FIG. 5.3 – Sélection de $i = \arg \min_{k=1,\dots,K} [\vec{f}(x) - \vec{\mu}_k]^T \Sigma_k^{-1} [\vec{f}(x) - \vec{\mu}_k]$ dans le cas où $P = 1$ et $K = 3$. Ici nous en déduisons $i = 1$ d'après (5.10). Le potentiel W final est la courbe représentée en pointillés.

5.2.3 Minimisation et algorithme

Comme au paragraphe 2.5 du chapitre 2, nous utilisons la méthode de *régularisation semi-quadratique* [Charbonnier *et al.*, 1997, Geman et Reynolds, 1992] permettant de surmonter la difficulté liée à la non linéarité des P EDP résultant des équations d'Euler-Lagrange pour la minimisation de (5.11). Les conditions que doit respecter la fonction φ pour pouvoir utiliser cette méthode sont les mêmes que celles présentées au sein du paragraphe 2.5 du chapitre 2.

En introduisant la variable auxiliaire $b : \mathbb{R}^2 \rightarrow \mathbb{R}$, nous ramenons la minimisation de (5.11) à celle de $J_\varepsilon^*(\vec{f}, b)$ par rapport au couple (\vec{f}, b) , avec :

$$J_\varepsilon^*(\vec{f}, b) = \int_{\Omega} |\vec{f}(x) - \vec{I}(x)|_{\mathbb{R}^P}^2 dx + \varepsilon \lambda^2 \int_{\Omega} [b(x) |\nabla \vec{f}(x)|_{\mathbb{R}^P}^2 + \psi(b(x))] dx + \frac{\eta^2}{\varepsilon} \int_{\Omega} W(\vec{f}(x)) dx, \quad (5.12)$$

avec ψ une fonction convexe en b dont l'expression analytique est déduite de celle de φ .

Par le biais des équations d'Euler-Lagrange, nous sommes alors conduits à résoudre alternativement les équations suivantes pour chaque valeur de ε :

$$\begin{cases} b = \frac{\varphi'(|\nabla \vec{f}|_{\mathbb{R}^P})}{2|\nabla \vec{f}|_{\mathbb{R}^P}} \in \mathbb{R} \text{ avec } \vec{f} \text{ fixé,} \\ \text{pour } p = 1, \dots, P : f^p - \lambda^2 \varepsilon \operatorname{div}(b \nabla f^p) + \frac{\eta^2}{2\varepsilon} W'_p(\vec{f}) = I^p, \text{ avec } b \text{ fixé.} \end{cases} \quad (5.13)$$

L'indice $p \in [1; P]$ désigne la $p^{\text{ième}}$ composante des vecteurs concernés dans les P EDP du schéma (5.13), et nous notons :

$$W'(\vec{f}) = [w'_1, \dots, w'_P]^T \quad \text{et} \quad W'_p(\vec{f}) = w'_p.$$

Pour chaque valeur de ε , nous trouvons une solution (\vec{f}, b) qui nous sert d'initialisation pour le système (5.13) lors de l'étape suivante de résolution pour une valeur plus faible du paramètre ε . En résolvant (5.13) pour chaque valeur de ε , nous obtenons une solution vectorielle \vec{f} dont chaque composante converge vers une solution de restauration et classification.

```

initialiser  $\vec{f}_0$ 
initialiser  $\varepsilon_0$ 
• répéter :
     $\varepsilon_{l+1} = \rho \varepsilon_l$  avec  $\rho < 1$ 
     $\vec{f}_{\varepsilon_{l+1}}^0 = \vec{f}_{\varepsilon_l}^{\min}$ 
    ◦ répéter :
         $f = \vec{f}_{\varepsilon_l}^q$ 
         $b = b_{\varepsilon_l}^q$ 
        incrémenter  $q$ 
        i) résoudre  $\min_b J_{\varepsilon_l}^*(f, b)$  avec  $f$  fixé :
            
$$b = \frac{\varphi'(|\nabla f|_{\mathbb{R}^P})}{2|\nabla f|_{\mathbb{R}^P}}$$

        ii) résoudre  $\min_f J_{\varepsilon_l}^*(f, b)$  avec  $b$  fixé :
            pour  $p = 1, \dots, P$   $f^p - \lambda^2 \varepsilon \operatorname{div}(b \nabla f^p) + \frac{\eta^2}{2\varepsilon} W'_p(f) = I^p$ 
        ◦ jusqu'à convergence sur  $\vec{f}_{\varepsilon_l}^q \rightarrow \vec{f}_{\varepsilon_l}^{\min}$ 
    • faire décroître  $\varepsilon_l$  jusqu'à convergence sur  $\vec{f}_{\varepsilon_l}^{\min}$ 

```

TAB. 5.2 – Algorithme associé à la minimisation de (5.12) par rapport à (\vec{f}, b) .

L'algorithme qui en découle est présenté dans le tableau 5.2. Dans ce tableau, Les P EDP de l'étape ii) issues de la minimisation de $J_{\varepsilon_l}^*$ par rapport à \vec{f} sont résolues par un algorithme de gradient conjugué. L'indice l est relatif à la boucle sur ε , alors que l'indice q est relatif aux étapes concernant les minimisations alternées (5.13). En pratique, l'initialisation \vec{f}_0 de \vec{f}_ε est telle que $\vec{f}_0 = \vec{I}$.

Les discrétisations adoptées dans l'algorithme présenté au sein du tableau 5.2 sont de même

nature que celles exposées dans l'annexe C.

Remarque concernant le calcul de $|\vec{\nabla}f|_{\mathbb{R}^P}$

Afin de combiner l'information des différentes bandes au niveau du calcul de $|\vec{\nabla}f|_{\mathbb{R}^P}$, nous utilisons des notions de géométrie riemannienne pour mesurer les variations de \vec{f} au lieu de la norme $L^2(\mathbb{R}^P)$ appliquée à $\vec{\nabla}f$. Cette approche a initialement été proposée par [Zenzo, 1986], on la retrouve également dans [Blomgren et Chan, 1998, Blomgren, 1998, Sapiro et Ringach, 1996, Torre et Poggio, 1986, Whitaker et Gerig, 1994]. Le principe de cette méthode consiste à extraire les directions de plus grandes variations d'une fonction à travers les caractéristiques de la forme fondamentale qui lui est associée.

Première forme fondamentale :

Soit $v : (x_1, x_2) \in \mathbb{R}^2 \rightarrow v(x_1, x_2) \in \mathbb{R}^P$ une fonction multidimensionnelle continue, dérivable.

Le *première forme fondamentale* associée à v s'écrit (voir par exemple [Kreyszig, 1959]) :

$$dv^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{\partial v}{\partial x_i} \cdot \frac{\partial v}{\partial x_j} dx_i dx_j \quad (5.14)$$

ou encore :

$$dv^2 = \begin{bmatrix} dx_1 \\ dx_2 \end{bmatrix}^T \underbrace{\begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}}_{[G]} \begin{bmatrix} dx_1 \\ dx_2 \end{bmatrix} \quad (5.15)$$

avec

$$g_{ij} = \sum_{p=1}^P \frac{\partial v^p}{\partial x_i} \frac{\partial v^p}{\partial x_j} \quad (5.16)$$

La première forme fondamentale dv^2 de v est une forme quadratique dont les coefficients g_{ij} permettent de trouver les directions ainsi que les valeurs extrémales des variations de v . La matrice $[G]$ est symétrique. Les valeurs propres λ_+ et λ_- sont données par :

$$\lambda^{\pm} = \frac{g_{11} + g_{22} \pm \sqrt{(g_{11} - g_{22})^2 + 4g_{12}^2}}{2}, \quad (5.17)$$

et les vecteurs propres associés par :

$$\theta_+ = \frac{1}{2} \arctan \left(\frac{2g_{12}}{g_{11} - g_{22}} \right), \quad \text{modulo } \pi$$

$$\theta_- = \theta_+ + \frac{\pi}{2}$$

θ_+ est la direction de plus grande variation de v en norme L^2 (avec λ_+ la valeur de cette variation), alors que θ_- est la direction de plus faible variation (avec λ_- la valeur de cette variation). On remarque que pour $P = 1$, c'est-à-dire dans le cas monospectral, nous avons $\lambda_+ = |\nabla v|^2$ et $\lambda_- = 0$.

Cette approche concernant le calcul du gradient, a été utilisée dans des schémas d'EDP pour la diffusion anisotrope permettant de lisser le bruit sur des images multispectrales, en combinant l'information des différentes bandes, et sans endommager les contours. Par exemple, un schéma de type :

$$\frac{\partial v}{\partial t} = h(\lambda_+, \lambda_-) \frac{\partial^2 v}{\partial \theta_-^2} \quad (5.18)$$

peut être considéré, où h est par exemple une fonction continue, décroissante de $(\lambda_+ - \lambda_-)$, un choix "naturel" étant $h(\lambda_+, \lambda_-) = \sqrt{\lambda_+ - \lambda_-}$.

L'EDP (5.18) permet de lisser les données dans les directions de plus faibles variations de l'intensité (si v représente l'intensité lumineuse des différentes bandes), voir par exemple [Blomgren et Chan, 1998, Blomgren, 1998, Sapiro et Ringach, 1996].

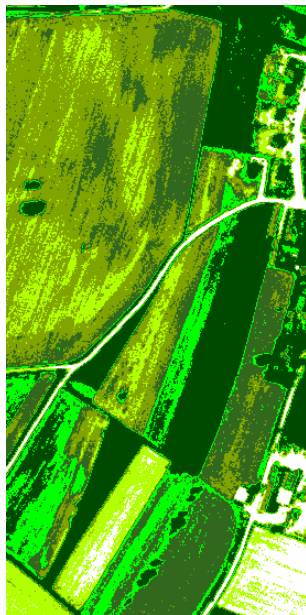
Notons que si le calcul de $|\nabla \vec{f}|_{\mathbb{R}^P}$ est remplacé par des calculs reposant sur des notions de géométrie riemannienne, c'est-à-dire en remplaçant $\varphi(|\nabla \vec{f}|_{\mathbb{R}^P})$ par $\varphi(\sqrt{\lambda_+ - \lambda_-})$ dans (5.11), l'expression des équations d'Euler-Lagrange n'est pas exactement celle donnée dans (5.13) car il faut alors dériver les valeurs propres extrémales par rapport aux composantes f^p . L'expression exacte de ces dérivées est donnée dans [Gadat, 1999]. Ici, nous utilisons les valeurs propres extrémales au niveau des EDP d'Euler-Lagrange pour calculer un gradient multispectral, et non au niveau de la fonctionnelle.

5.2.4 Exemple : imagerie couleur

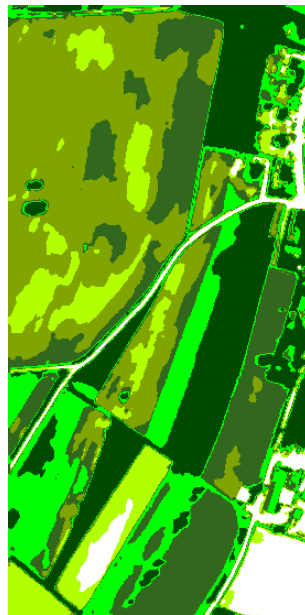
A titre indicatif, nous montrons un exemple de résultat obtenu sur une image couleur aérienne, à 0.5 mètre de résolution, fournie par l'Institut Géographique National (IGN). En effet, l'approche que nous avons adoptée permet d'effectuer une classification sur des images couleurs, les trois canaux RVB (rouge, vert et bleu) jouant le rôle de trois bandes spectrales. Nous avons au préalable estimé le nombre de classes ainsi que leurs paramètres à l'aide de l'algorithme de *fuzzy Cmeans* avec terme d'entropie que nous avons présenté en début de chapitre. Le nombre initial de classes Cl est de 9 et au bout de 10 itérations de l'algorithme de *fuzzy Cmeans* modifié, nous obtenons un nombre optimal de classes K égal à 6. Le poids α de la contrainte d'entropie dans (5.1) est de 5.10^6 . Nous présentons



a) données aériennes couleur



b) fuzzy Cmeans modifié



c) classification obtenue

FIG. 5.4 – Résultats de classification sur des données aériennes en couleur (fig. a) obtenus par l'algorithme de *fuzzy Cmeans* modifié (fig. b) et par l'algorithme présenté dans le tableau 5.2.

les données initiales sur la figure 5.4 (fig. a) et le premier résultat de classification donné par l'algorithme de *fuzzy Cmeans* (fig. b). Le résultat de classification obtenu par le modèle multispectral défini par (5.12) est présenté sur la figure 5.4 (fig. c), après 60 itérations sur la décroissance de ε (cf. algorithme dans le tableau 5.2). La valeur initiale de \vec{f} est celle des données \vec{I} , nous n'utilisons donc pas le résultat préliminaire de classification fourni par l'algorithme du *fuzzy Cmeans* modifié (fig. b).

5.3 Extension du deuxième modèle

5.3.1 Introduction

Le modèle de classification par régions et contours actifs présenté au chapitre 3 s'étend au cas de données $\vec{I} : \mathbb{R}^2 \rightarrow \mathbb{R}^P$. Les propriétés de régularité de la solution étant les mêmes que dans le cas monospectral, et compte tenu de l'expression du critère monospectral, seul le terme d'attache aux données sera affecté par un passage au multispectral. C'est à travers ce terme que sera combinée l'information des différentes bandes spectrales, les autres termes jouant sur la partition du support Ω qui est unique pour toutes les bandes.

5.3.2 Fonctionnelle multispectrale

Rappelons l'expression de la fonctionnelle (3.35) que nous avons définie dans le cas monospectral :

$$\begin{aligned}
 F_\alpha(\Phi_1, \dots, \Phi_K) = & \underbrace{\sum_{i=1}^K e_i \int_{\Omega} H_\alpha(\Phi_i) \frac{(I(x) - \mu_i)^2}{\sigma_i^2} dx}_{\text{attache aux données}} \\
 & + \underbrace{\sum_{i=1}^K \gamma_i \int_{\Omega} g(|\nabla I|) \delta_\alpha(\Phi_i) |\nabla \Phi_i| dx}_{\text{minimisation de la longueur des contours}} \\
 & + \underbrace{\frac{\lambda}{2} \int_{\Omega} \left(\sum_{i=1}^K H_\alpha(\Phi_i) - 1 \right)^2 dx}_{\text{contrainte de partition}}.
 \end{aligned}$$

Lorsque $\vec{I} \in L^2(\Omega, \mathbb{R}^P)$, nous cherchons une partition des données vectorielles composées d'un ensemble de K régions Ω_i séparées par des interfaces de longueurs minimales Γ_i . Les conditions A et C du tableau 3.1 imposées sur la solution de partition dans le cas monospectral restent identiques dans le cas multispectral : nous voulons une et une seule étiquette par pixels et nous souhaitons obtenir des frontières régularisées. Seule la condition

B doit prendre en compte le caractère vectoriel des données et combiner l'information des P bandes à travers la matrice de variance-covariance Σ_i définie pour chaque classe C_i . Chaque classe C_i est caractérisée par une distribution gaussienne multidimensionnelle de l'intensité \vec{I} sur l'ensemble des régions Ω_i :

$$\Omega_i = \{x \in \Omega / \vec{I}(x) \sim \mathcal{N}_{\mathbb{R}^P}(\vec{\mu}_i, \Sigma_i)\}$$

En approchant l'ensemble Ω_i à l'aide de la fonction d'ensemble de niveaux $\Phi_i : \Omega \rightarrow \mathbb{R}$, définie comme une fonction de distance signée, à laquelle est appliquée une approximation de Heaviside (3.20), nous définissons l'attache aux données par :

$$F_{\alpha}^{\text{DONNÉES}}(\Phi_1, \dots, \Phi_K) = \sum_{i=1}^K e_i \int_{\Omega} H_{\alpha}(\Phi_i) [\vec{I} - \vec{\mu}_i]^T \Sigma_i^{-1} [\vec{I} - \vec{\mu}_i] dx. \quad (5.19)$$

Une matrice Σ_i diagonale, signifie que les bandes sont complètement décorrélées concernant la classe C_i .

Comme pour le cas monospectral (cf. paragraphe 3.6), remarquons que dans le cas où les classes ont les mêmes moyennes et des variances différentes, le terme d'attache aux données défini dans (5.19) n'est pas discriminant. On peut alors envisager un terme supplémentaire tenant compte du fait que l'intensité \vec{I} suit une loi de distribution gaussienne sur les classes :

$$F_{\alpha}^{\text{DONNÉES}}(\Phi_1, \dots, \Phi_K) = \sum_{i=1}^K e_i \int_{\Omega} H_{\alpha}(\Phi_i) \left\{ [\vec{I} - \vec{\mu}_i]^T \Sigma_i^{-1} [\vec{I} - \vec{\mu}_i] + \log \left(\sqrt{(2\pi)^P \det(\Sigma_i)} \right) \right\} dx. \quad (5.20)$$

Nous avons négligé le terme logarithmique par la suite car son influence est très faible dans les applications traitées.

Les contraintes de partition et sur la longueur des interfaces restant inchangées dans le cas multispectral, nous proposons de minimiser par rapport aux fonctions $\{\Phi_i\}_i$ le critère

multispectral suivant :

$$\begin{aligned}
 F_\alpha(\Phi_1, \dots, \Phi_K) = & \underbrace{\sum_{i=1}^K e_i \int_{\Omega} H_\alpha(\Phi_i) [\vec{I} - \vec{\mu}_i]^T \Sigma_i^{-1} [\vec{I} - \vec{\mu}_i] dx}_{\text{attache aux données multispectrales}} \\
 & + \underbrace{\sum_{i=1}^K \gamma_i \int_{\Omega} g(|\nabla I|) \delta_\alpha(\Phi_i) |\nabla \Phi_i| dx}_{\text{minimisation de la longueur des contours}} \\
 & + \underbrace{\frac{\lambda}{2} \int_{\Omega} \left(\sum_{i=1}^K H_\alpha(\Phi_i) - 1 \right)^2 dx}_{\text{contrainte de partition}}. \tag{5.21}
 \end{aligned}$$

Les différents termes du critère F_α doivent permettre, lors de la minimisation par rapport aux fonctions Φ_i , d'obtenir une partition régulière des P bandes tenant compte des caractéristiques gaussiennes des classes.

5.3.3 Schéma algorithmique

La minimisation de la fonctionnelle (5.21) repose sur l'annulation des K dérivées de F_α rapport aux fonctions Φ_i . Nous plongeons les EDP d'Euler-Lagrange dans un schéma dynamique où t représente le paramètre de temps, et nous obtenons les K schémas suivants, pour $i = 1, \dots, K$:

$$\begin{aligned}
 \Phi_i^{t+1} = \Phi_i^t - dt \Big\{ & \delta_\alpha(\Phi_i^t) \left[e_i [\vec{I} - \vec{\mu}_i]^T \Sigma_i^{-1} [\vec{I} - \vec{\mu}_i] - \gamma_i g(|\nabla I|) \operatorname{div} \left(\frac{\nabla \Phi_i^t}{|\nabla \Phi_i^t|} \right) - \gamma_i \frac{\nabla g \nabla \Phi_i^t}{|\nabla \Phi_i^t|} \right. \\
 & \left. + \lambda \left(\sum_{j=1}^K H_\alpha(\Phi_j^t) - 1 \right) \right] \Big\}. \tag{5.22}
 \end{aligned}$$

Ces K EDP sont couplées au niveau du terme de partition $\left(\sum_{j=1}^K H_\alpha(\Phi_j^t) - 1 \right)$. L'évolution de chaque fonction d'ensemble de niveaux Φ_i au cours du temps, et par là même celle des régions Ω_i et des contours associés Γ_i , est guidée par un jeu de forces contraignant la solution de partition finale à respecter les propriétés de régularité souhaitées. L'algorithme qui en découle est présenté dans le tableau 5.3. Nous souhaitons que les fonctions Φ_i restent des fonctions de distance signée au cours de leur évolution à travers le schéma (5.22). Comme dans le cas monospectral, afin de préserver la contrainte $|\nabla \Phi_i| = 1$, nous devons régulariser fréquemment¹ chaque fonction Φ_i au cours du temps par l'EDP (3.7) proposée

¹voir le paragraphe 3.1 concernant la perte de régularité des fonctions de distance au cours du temps, lors de la mise à jour des fonctions Φ_i par l'EDP (5.22).

par [Sussman *et al.*, 1994].

Les discrétisations adoptées dans l'algorithme présenté dans le tableau 5.3 sont de même

<pre> 0 - initialiser Φ_i^0 pour $i = 1...K$ 1 - $t \leftarrow t + 1$ 2 - pour $i = 1...K$ résoudre les K EDP couplées de (5.22) 3 - toutes les n itérations de (5.22), régulariser les fonctions Φ_i à l'aide de l'EDP (3.7) 4 - aller à l'étape 1 pour incrémenter le temps t </pre>

TAB. 5.3 – Algorithme associé au schéma (5.22).

nature que celles exposées en annexe F.

5.3.4 Exemple : imagerie couleur

A titre indicatif, nous montrons un exemple de résultat obtenu sur l'image couleur aérienne déjà présentée sur la figure 5.4. Nous présentons sur la figure 5.5 les données initiales (fig.a) et le premier résultat de classification donné par l'algorithme de *fuzzy Cmeans* (fig. b), dont les paramètres choisis (valeur de α , nombre initial de classes, nombre d'itérations) sont les mêmes que ceux de la figure 5.4.

Nous utilisons deux types d'initialisation des fonctions Φ_i pour l'algorithme donné dans le tableau 5.3 : une initialisation automatique par des fonctions Φ_i circulaires dans des fenêtres de taille 3×3 (cf. chapitre 3, figure 3.7), donnant le résultat de classification présenté sur la figure 5.5 (fig. c), et une initialisation des Φ_i par le résultat de classification obtenu à l'aide de l'algorithme de *fuzzy Cmeans* (fig. b), aboutissant au résultat final de classification présenté sur la figure d. Nous constatons sur la figure 5.5 la régularisation supplémentaire obtenue par le modèle de régions/contours actifs que nous proposons (fig. c et d) par rapport au résultat de classification du *fuzzy Cmeans* (fig. b), indépendamment du type d'initialisation choisi. Les classes que nous obtenons sont bien plus homogènes et les frontières finales sont relativement lisses. Nous avons délibérément choisi une forte régularisation (poids élevé du terme de contrainte sur la longueur des contours dans (5.21)) pour bien insister sur le pouvoir régularisant de ce modèle.

Nous constatons une légère différence sur les résultats de classifications selon le type d'initialisation (fig.c et fig. d). Il est difficile d'en déduire le meilleur type d'initialisation en visualisant les résultats obtenus. Cependant, une initialisation des Φ_i par une classification erronée de l'algorithme de *fuzzy Cmeans* pourrait conduire à une mauvaise classification

par régions et contours actifs. Dans ce cas, il semble préférable d'utiliser uniquement l'algorithme de *fuzzy Cmeans* modifié pour l'estimation du nombre et des paramètres des classes, par précaution. Le problème ne se pose pas pour le paragraphe suivant, car nous n'utilisons par le modèle de *fuzzy Cmeans* modifié puisque le nombre de classes ainsi que leurs paramètres ont été estimés par un autre algorithme dont les résultats nous ont été fournis par le projet VISTA de l'IRISA.

Nous présentons sur la figure 5.6 l'évolution des régions et de leurs contours (ENZ) au cours des itérations de l'algorithme du tableau 5.3 pour une initialisation automatique des Φ_i , correspondant au résultat de classification c) de la figure 5.5. Sur la figure 5.7, nous montrons l'évolution des régions et de leurs contours pour une initialisation des Φ_i par le résultat de classification donné par l'algorithme de *fuzzy Cmeans*, et aboutissant au résultat de classification d) de la figure 5.5.

5.4 Conclusion

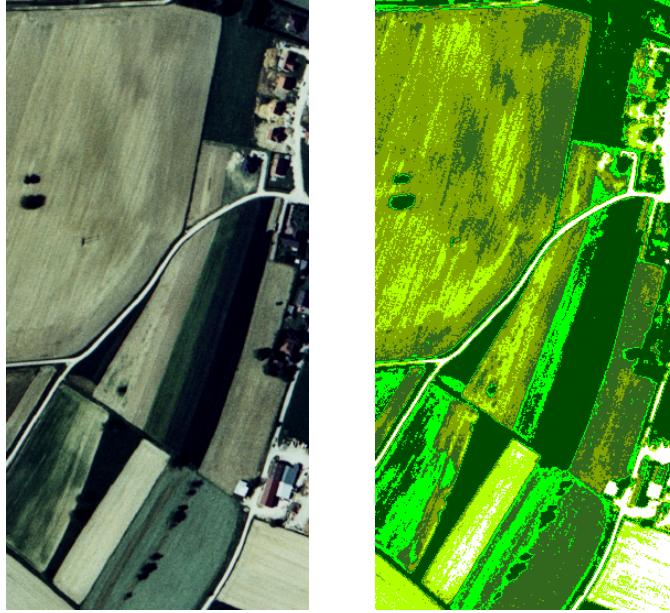
Nous avons étendu les modèles variationnels de classification développés au sein des chapitres 2 et 3 au cas des données multispectrales. Cette extension intervient à différents niveaux pour ces deux modèles.

L'extension du modèle combinant restauration et classification concerne tous les termes de la fonctionnelle définie dans le cas monospectral, dont le terme de contrainte de niveaux (potentiel W) pour lequel l'extension au cas multispectral est assez délicate à effectuer. Les résultats théoriques issus de la théorie de Van de Waals-Cahn-Hilliard restent valables pour le cas multispectral.

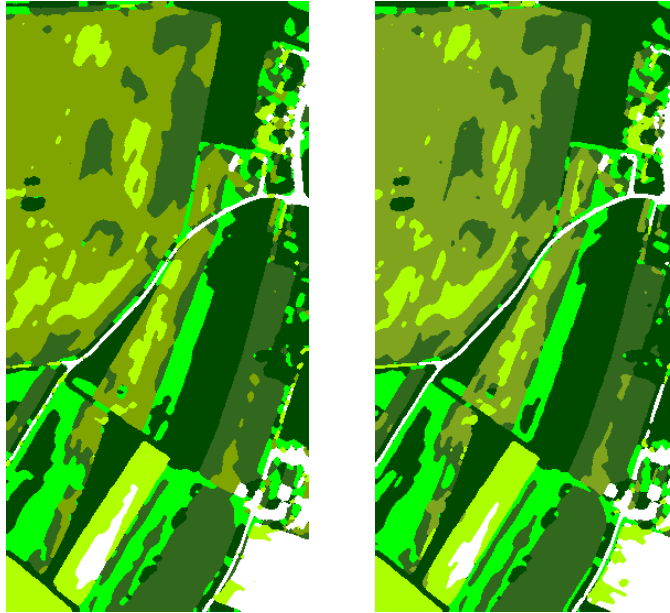
L'extension du modèle dynamique est quasi immédiate, puisqu'elle ne concerne qu'un seul terme de la fonctionnelle définie dans le cas monospectral, le terme d'attache aux données, la seule difficulté consistant à calculer une norme dans un espace multi-dimensionnel.

Nous avons testé ces deux modèles sur des images aériennes en couleur, dont les canaux RVB ont été assimilés à différentes bandes spectrales, comme les bandes XS1, XS2 et XS3 des satellites SPOT.

Dans le chapitre suivant, nous traitons une application à partir d'une scène multispectrale SPOT 2 pour laquelle nous disposons d'une vérité terrain ainsi que de différents résultats de classification fournis par le projet VISTA de l'IRISA.



a) données aériennes couleur b) fuzzy Cmeans modifié



c) classif. pour init. 1 d) classif. pour init. 2

FIG. 5.5 – Résultats de classification sur des données aériennes en couleur (fig. a) obtenus par l'algorithme de *fuzzy Cmeans* modifié (fig. b) et par modèle de régions/contours actifs pour 2 types d'initialisation ; init. 1 : initialisation des Φ_i dans des fenêtres 3×3 (fig. c) ; init.2 : pour des Φ_i initialisés par le résultat de classification donné par l'algorithme de *fuzzy Cmeans* modifié (fig. d).

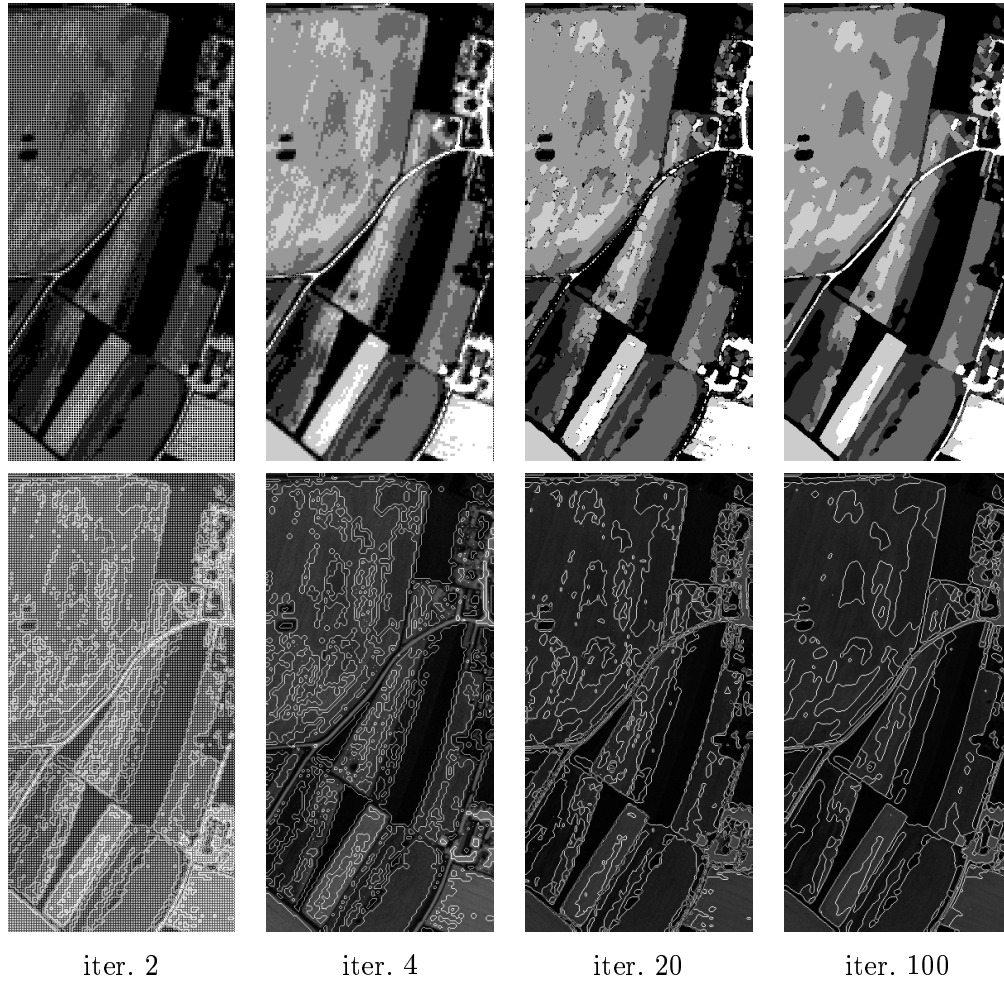


FIG. 5.6 – Image couleur de l'IGN : évolution des régions (en haut) et de leurs contours (en bas) au cours des itérations pour une initialisation des Φ_i dans des fenêtres 3×3 .

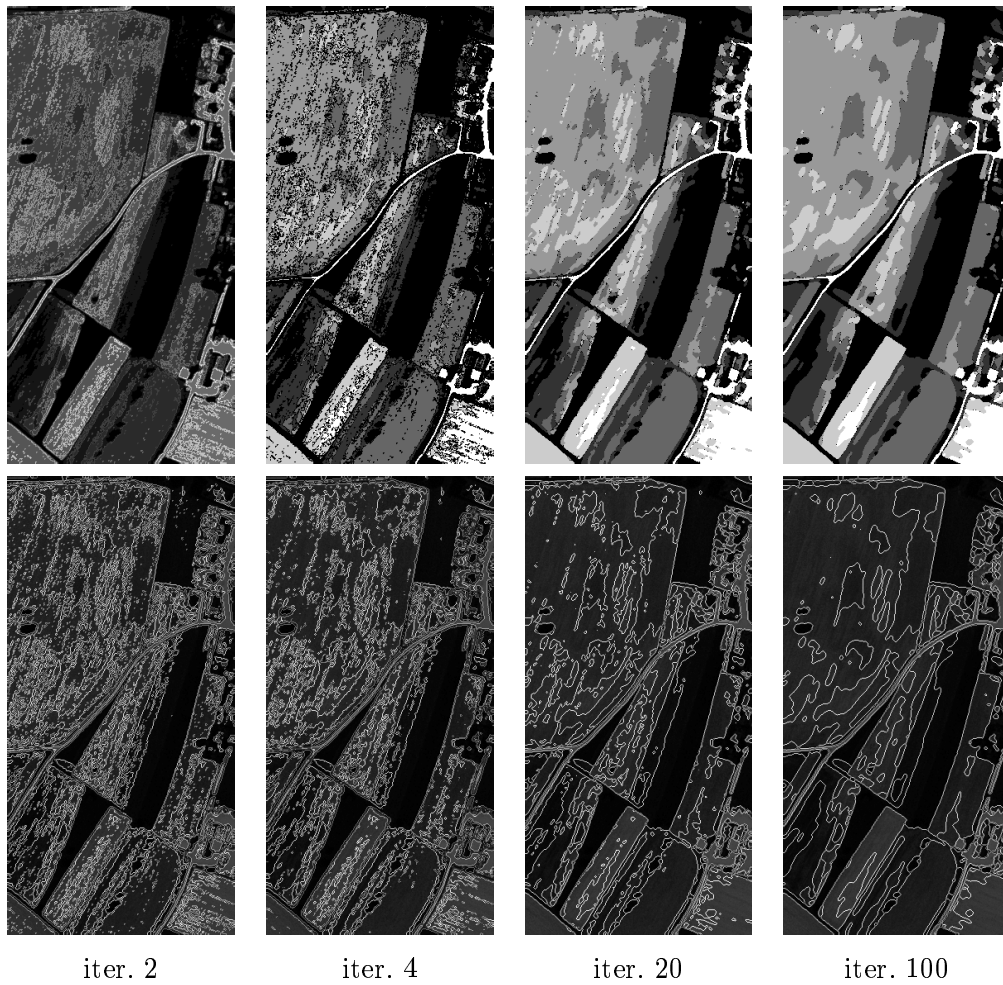
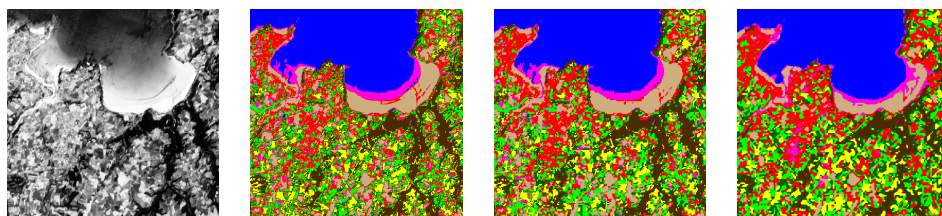


FIG. 5.7 – Image couleur de l'IGN : évolution des régions (en haut) et de leurs contours (en bas) au cours des itérations pour une initialisation des Φ_i par le résultat de classification obtenu par *fuzzy Cmeans* modifié.

Chapitre 6

Une application sur des données multispectrales SPOT



Où nous traitons une application réelle sur une scène SPOT en mode XS pour laquelle nous disposons d'une vérité terrain. Nous effectuons une classification de cette scène à l'aide des versions multispectrales des modèles de classification présentées au chapitre précédent : classification avec restauration et classification par modèle dynamique. Nous comparons ces deux modèles à d'autres approches dont un modèle stochastique hiérarchique, récemment développé à l'IRISA au sein du projet VISTA.

6.1 Présentation des données et des modèles de comparaison

6.1.1 Présentation des données

Les données multispectrales SPOT 2 de la baie de Lannion présentées sur la figure 6.2 ont fait l'objet d'une étude approfondie dans la thèse d'Annabelle Chardin [Chardin, 2000] menée à l'IRISA dans le projet VISTA. Cette étude entre dans le cadre d'une collaboration entre plusieurs laboratoires de la région Bretagne impliqués dans la télédétection et regroupés au sein du GSTB (Groupement Scientifique pour la Télédétection en Bretagne). Ces données permettent de mesurer l'impact de la culture intensive sur l'évolution de l'utilisation des sols dans cette baie. Une étude de l'occupation des sols à partir d'images prises à différentes époques (images SPOT 2 en mode XS prises en hiver 1996 et été 1997) a permis d'analyser cette évolution. A. Chardin a développé un algorithme de classification fondé sur des champs de Markov hiérarchiques que nous présentons par la suite. Les résultats qui ont été obtenus dans le cadre de sa thèse ont été comparés à deux autres modèles, le maximum de vraisemblance MV (cf. chapitre 1 paragraphe 1.3.2) ainsi qu'à l'ICM niveau 0 qui correspond au premier niveau du modèle hiérarchique mentionné ci-dessus. De plus, les géographes du COSTEL (Climat, Occupation des Sols par TELédétection) de l'Université de Rennes II ont fourni une vérité terrain permettant de valider ces résultats. Tous ces résultats ont été gracieusement mis à notre disposition par le projet VISTA de l'IRISA Rennes que nous tenons à remercier, et les données SPOT de la scène d'été 1997 ont été acquises auprès de la société SPOTIMAGE. Notre objectif n'est pas de réaliser une étude approfondie de l'utilisation des sols, comme cela a pu être le cas dans [Chardin, 2000], mais de pouvoir comparer nos résultats de classification avec d'autres approches sur des données satellitaires multispectrales pour lesquelles nous disposons d'une vérité de terrain établie par des experts géographes.

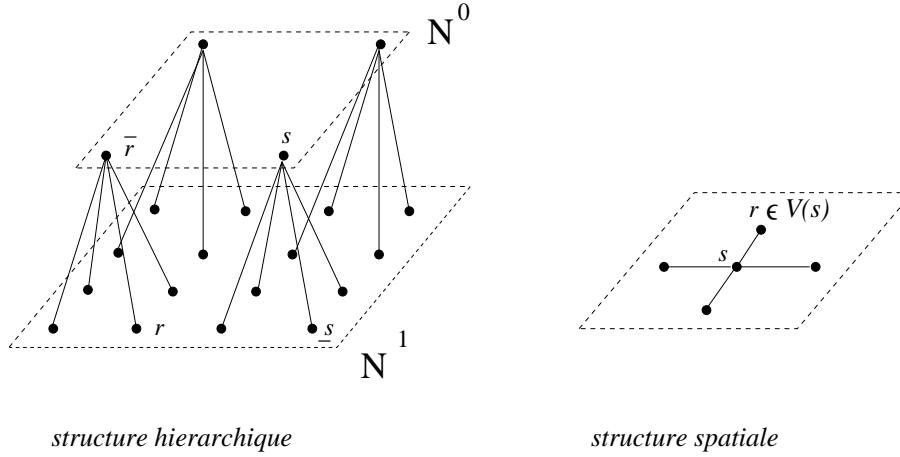


FIG. 6.1 – Structure de l'arbre tronqué.

6.1.2 Modèle markovien hiérarchique sur arbre tronqué

L'approche adoptée dans [Chardin, 2000] entre dans le cadre des modèles markoviens hiérarchiques. Le lecteur intéressé par ces approches et algorithmes associés pourra trouver de plus amples informations sur le sujet dans la synthèse [Graffigne *et al.*, 1995] ou dans les travaux [Bouman et Shapiro, 1994, Heitz *et al.*, 1994, Kato *et al.*, 1999, Kato, 1994, Laferté, 1996, Pérez, 1993, Pérez et Heitz, 1996]. L'originalité du modèle hiérarchique proposé dans [Chardin, 2000] repose sur l'introduction d'une structure hiérarchique permettant de combiner des a priori spatiaux et hiérarchiques : il s'agit d'un quadrarbre tronqué dont les noeuds du niveau le plus grossier supportent un champ de Markov spatial sur une grille.

Nous utilisons les mêmes notations que dans le paragraphe 1.3.2 du chapitre 1, en particulier, dans le cadre stochastique, on cherche à estimer un étiquetage représenté par la variable L , étant données les observations représentées par la variable O . Nous ne détaillerons pas ce modèle, mais nous nous contenterons d'expliciter les points essentiels à sa compréhension générale. Tous les détails relatifs à cette approche sont clairement présentés dans [Chardin, 2000].

Considérons une structure pyramidale de $M + 1$ niveaux de résolution. Soit N^m l'ensemble des noeuds du niveau m , avec $m \in [0; M]$, où N^M est le niveau de résolution le plus fin et N^0 le plus grossier. L'arbre tronqué est un quadrarbre dont la racine N^0 n'est pas un pixel, mais une grille spatiale rectangulaire munie d'un système de voisinage du premier ordre et dont chaque noeud est la racine d'un quadrarbre. On note \mathcal{S} l'ensemble des

Passe montante

- Feuilles s , $s \in N^M$:

$$V_s(L_{\bar{s}}) = \min_{L_s} [h_s(L_s, L_{\bar{s}}) + g_s(L_s, O_s)]$$

$$L_s^*(L_{\bar{s}}) = \arg \min_{L_s} [h_s(L_s, L_{\bar{s}}) + g_s(L_s, O_s)]$$

- Récurrence pour $m = M - 1, \dots, 1$ et $s \in N^m$:

$$V_s(L_{\bar{s}}) = \min_{L_s} [h_s(L_s, L_{\bar{s}}) + g_s(L_s, O_s) + \sum_{r \in \underline{s}} V_r(L_s)]$$

$$L_s^*(L_{\bar{s}}) = \arg \min_{L_s} [h_s(L_s, L_{\bar{s}}) + g_s(L_s, O_s) + \sum_{r \in \underline{s}} V_r(L_s)]$$

ICM sur la grille grossière N^0

- Initialisation :

$$\forall s \in N^0, L_s^{(0)} = \arg \min_{\lambda} [g_s(\lambda, O_s) + \sum_{r \in \underline{s}} V_r(\lambda)]$$

- Effectuer plusieurs mises à jour de tous les sites de N^0 ,
en faisant décroître l'énergie

$$\sum_{\langle r, s \rangle \in N^0} f_{r,s}(L_r, L_s) + \sum_{s \in N^0} [g_s(L_s, O_s) + \sum_{r \in \underline{s}} V_r(L_s)]$$

pour obtenir \hat{L}_s , $\forall s \in N^0$

Passe descendante

pour $m = 1, \dots, M$, $s \in N^m$

$$\hat{L}_s = L_s^*(\hat{L}_{\bar{s}})$$

TAB. 6.1 – Algorithme de minimisation de (6.1), extrait de [Chardin, 2000].

enfants de s , \bar{r} le parent du site r (voir l'illustration sur la figure 6.1) et \underline{s} l'ensemble des sites formant l'arbre de racine s .

Dans le cadre de l'estimateur du MAP (cf. chapitre 1, paragraphe 1.3.2), l'estimé \hat{L} de l'étiquetage L est obtenu par :

$$\hat{L} = \arg \min_L E(L), \quad \text{où } E \text{ est l'énergie globale définie par :}$$

$$\begin{aligned} E(L) = & \sum_{s \in N^0} g_s(L_s, O_s) + \sum_{\langle r, s \rangle \in N^0} \underbrace{f_{r,s}(L_r, L_s)}_{\text{Potts spatial}} \\ & + \sum_{s \in N \setminus N^0} \left[\underbrace{h_s(L_s, L_{\bar{s}})}_{\text{Potts hiérarchique}} + g_s(L_s, O_s) \right] \end{aligned} \quad (6.1)$$

avec g_s une fonction quadratique prenant en compte la répartition gaussienne de l'intensité lumineuse dans les classes (i.e. attache aux données). Les deux termes d'a priori $f_{r,s}$ et h_s sont des modèles de Potts (cf. définition du modèle de Potts au chapitre 1, paragraphe 1.3.2), $f_{r,s}$ étant un a priori de Potts *spatial* (en 4-connexité), alors que h_s est un a priori de Potts *hiérarchique*.

On pose :

$$V_s(L_{\bar{s}}) = \min_{L_{\bar{s}}} \left[\sum_{r \in \bar{s}} [h_r(L_r, L_{\bar{r}}) + g_r(L_r, O_r)] \right] \quad (6.2)$$

et l'énergie E peut alors s'écrire :

$$\min_L E(L) = \min_{L_s, s \in N_0} \left\{ \sum_{\langle r, s \rangle \in N^0} f_{r,s}(L_r, L_s) + \sum_{s \in N^0} \left[g_s(L_s, O_s) + \sum_{r \in \bar{s}} V_r(L_s) \right] \right\}. \quad (6.3)$$

L'algorithme hiérarchique d'optimisation de (6.3) est présenté dans le tableau 6.1. Il se décompose en trois étapes : une passe montante sur les sous-arbres dont la racine est un site de N^0 , un algorithme ICM sur la grille grossière N^0 et une passe descendante sur les sous-arbres parcourus lors de la passe montante.

6.2 Résultats et comparaisons

6.2.1 Présentation des résultats

Nous avons conduit des expérimentations sur des images SPOT 2 en mode XS de la baie de Lannion. Ces images datent du 15 Août 1997, et nous avons choisi de les acquérir pour pouvoir comparer nos résultats de classification avec ceux fournis par le projet VISTA. Les images originales issues des 3 bandes spectrales ont une taille de 1480×1024 pixels, et nous avons mené nos expérimentations sur une portion de la scène de taille 400×400 pixels. Cette zone a été choisie pour la diversité morphologique et thématique des objets qui la composent. L'intérêt de cette zone provient également du fait que la densité d'échantillons de la vérité terrain fournie par le COSTEL est importante, comparée au reste de la scène globale. Nous présentons sur la figure 6.2 les 3 scènes XS1, XS2 et XS3 considérées.

Par la suite, nous désignerons les différents résultats de classification par :

- **MV**, pour le Maximum de Vraisemblance (cf. paragraphes 1.3.2 et 4.2)
- **ICM-N0**, pour le modèle (6.1) proposé par A. Chardin, mais sans aspect hiérarchique, c'est-à-dire pour $N = 0$
- **H-MAP**, pour le modèle hiérarchique sur arbre tronqué proposé par A. Chardin (6.1), avec pour niveau de résolution le plus fin $N = 3$
- **M1X**, pour la version multispectrale du premier modèle que nous avons proposé (cf. algorithme dans le tableau 5.2)

- **M2X**, pour la version multispectrale du second modèle que nous avons proposé (cf. algorithme dans le tableau 5.3)

Les résultats de classification MV, ICM-N0 et H-MAP proviennent des travaux de [Chardin, 2000]. Notons que ces trois algorithmes ont été expérimentés sur les scènes complètes de taille 1480×1024 pixels et sur une machine différente de la nôtre, nous ne comparerons donc pas les temps de calcul. Concernant les résultats H-MAP, la taille initiale de la grille de niveau $N = 0$ est de 185×128 pixels.

Le nombre de classes est fixé par les thématiciens et géographes du laboratoire COSTEL, et la légende des thèmes d'occupation des sols est représentée sur la figure 6.4. Les géographes ont sélectionné des zones différentes de l'image, sous forme de petites parcelles rectangulaires, pour l'estimation des paramètres des 8 classes d'une part (parcelles d'apprentissage), et pour valider les résultats de classification d'autre part (parcelles de vérité terrain). La figure 6.3 présente ces parcelles sur des images de taille 400×400 pixels extraites des images originales dont la taille est de 1480×1024 pixels. Notons au passage que les parcelles d'apprentissage et de validation ne sont pas les mêmes. Les paramètres des classes que nous avons utilisés ont été estimés sur les parcelles d'apprentissage de l'image complète 1480×1024 et sont présentés dans [Chardin, 2000], alors que nous effectuons la validation des résultats obtenus par les différents modèles à partir des extraits 400×400 de la vérité terrain.

Concernant l'estimation des paramètres, le nombre d'échantillons diffère selon les classes. En particulier pour la classe "légume" (classe 7) il n'y a que 38 pixels d'échantillons pour l'apprentissage : les résultats pour cette classe sont par conséquent peu fiables, voire insignifiants, d'autant que, dans la portion d'image 400×400 que nous avons choisi de traiter, la vérité terrain ne comporte que 5 pixels de validation pour cette classe. Il en est quasiment de même pour la classe 4 ("bois et landes") puisque la vérité terrain ne comporte que 14 pixels dans la portion 400×400 que nous avons sélectionnée. La vérité terrain contient au total 1466 pixels, permettant de localiser et de compter le nombre de pixels bien/mal classés sur les petites zones rectangulaires représentées sur la figure 6.3. Le nombre de pixels par classe pour la vérité terrain est comptabilisé sur la dernière ligne du tableau 6.3.

Nous présentons dans un premier temps le résultat de classification obtenu par MV sur la figure 6.4, c'est-à-dire sans aucune régularisation sur la solution cherchée. Nous constatons que la classification comporte un grand nombre de petites régions et que les classes sont peu homogènes.

Sur la figure 6.5, nous présentons les quatre résultats de classification obtenus par les mo-

dèles ICM-N0, H-MAP, M1X et M2X. Pour le modèle M1X, nous avons effectué 40 itérations sur la décroissance de ε (cf. algorithme dans le tableau 5.2 du chapitre 5). Concernant le modèle M2X, nous avons effectué 60 itérations pour ce qui est de l'évolution dynamique des régions et de leurs contours (cf. algorithme dans le tableau 5.3 du chapitre 5), et nous avons initialisé les régions associées aux fonctions Φ_i ($i = 1...8$) dans des fenêtres de taille 5×5 (cf. procédure d'initialisation automatique des fonctions Φ_i sur la figure 3.7 du chapitre 3).

modèle	nombre d'erreurs	pourcentage de réussite
MV	512	65.07%
ICM-N0	498	66.03%
H-MAP	433	70.46%
M1X	421	71.28%
M2X	440	69.99%

TAB. 6.2 – Nombre d'erreurs commises et pourcentages de réussite pour les différents modèles de classification

6.2.2 Commentaires

A première vue, nous pouvons constater sur la figure 6.5 que la régularité de la solution est différente selon les modèles de classification. Le modèle M2X est celui qui fournit le plus de régularisation dans le sens où les régions obtenues sont très homogènes avec des frontières relativement lisses. Cette régularité est le résultat conjoint du terme de contrainte sur la longueur des contours de la fonctionnelle proposée (5.21) et surtout de la procédure automatique d'initialisation des fonctions Φ_i dans des fenêtres 5×5 , limitant la résolution spatiale du modèle. La procédure automatique semble avoir plus d'impacts sur la morphologie des régions que le terme de minimisation de longueur des contours dans le sens où les régions obtenues n'ont pas nécessairement une faible courbure puisque nous trouvons un certain nombre de "coins".

Nous pouvons constater que les résultats des modèles ICM-N0 et H-MAP conduisent à de nombreuses régions dont les contours sont horizontaux ou verticaux. Ce phénomène a sans doute pour origine la 4-connexité des modèles de Potts utilisés dans ces deux modèles.

Nous remarquons que la classe 7 ("légumes"), dont nous avons mentionné les problèmes en termes de validité, est gérée de façon très similaire par les modèles ICM-N0, H-MAP et M2X, alors que le modèle M1X conduit à des résultats assez différents. Même si cette classe n'est pas significative pour la validation des résultats, il est intéressant de noter le comportement différent du modèle M1X par rapport aux trois autres modèles, dont le terme de



canal XS1

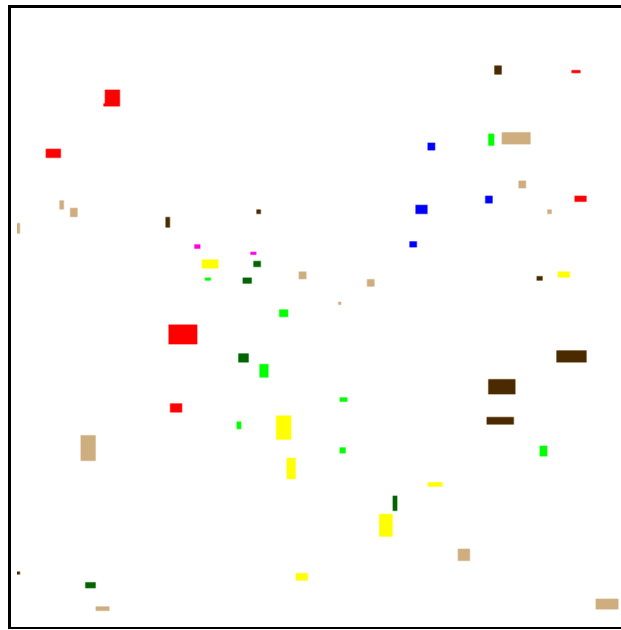


canal XS2

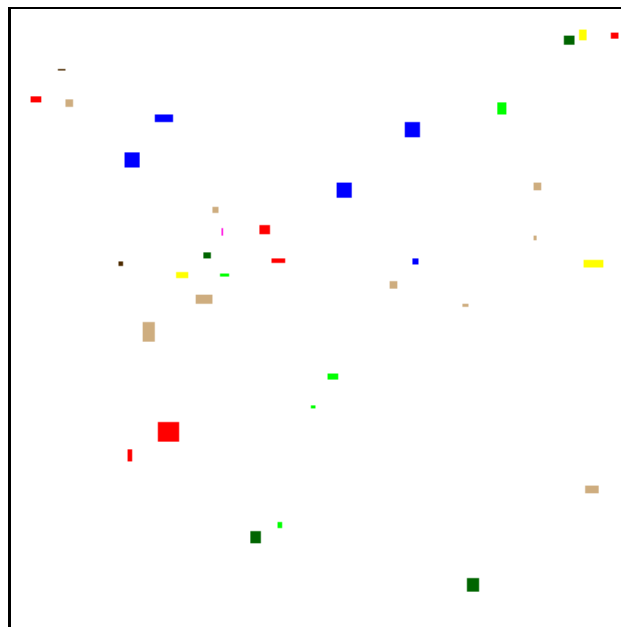


canal XS3

FIG. 6.2 – Données multispectrales SPOT 2 de la baie de Lannion (Août 1997) après égalisation de l'histogramme.



zones d'apprentissage



zones de validation

FIG. 6.3 – Zones d'apprentissage (2719 pixels) dans lesquelles les paramètres des 8 classes ont été estimés et zones servant de validation (1466 pixels) pour les résultats de classification obtenus (vérité terrain).

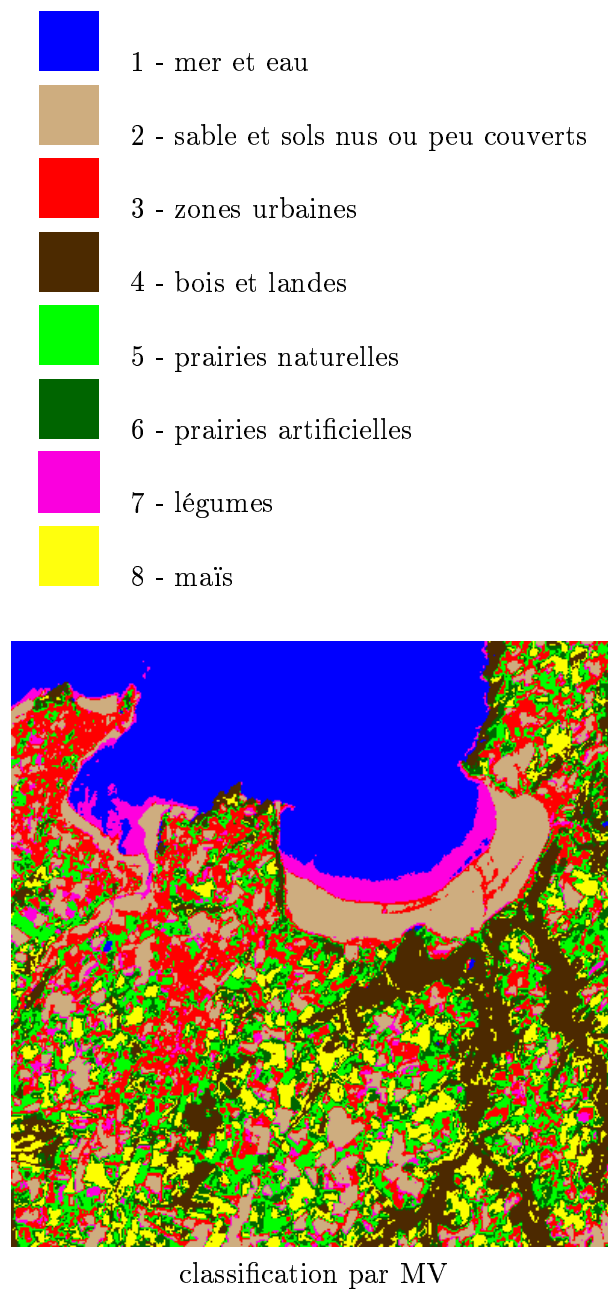


FIG. 6.4 – Légende des thèmes d'occupation des sols et résultat de classification pour le modèle MV.

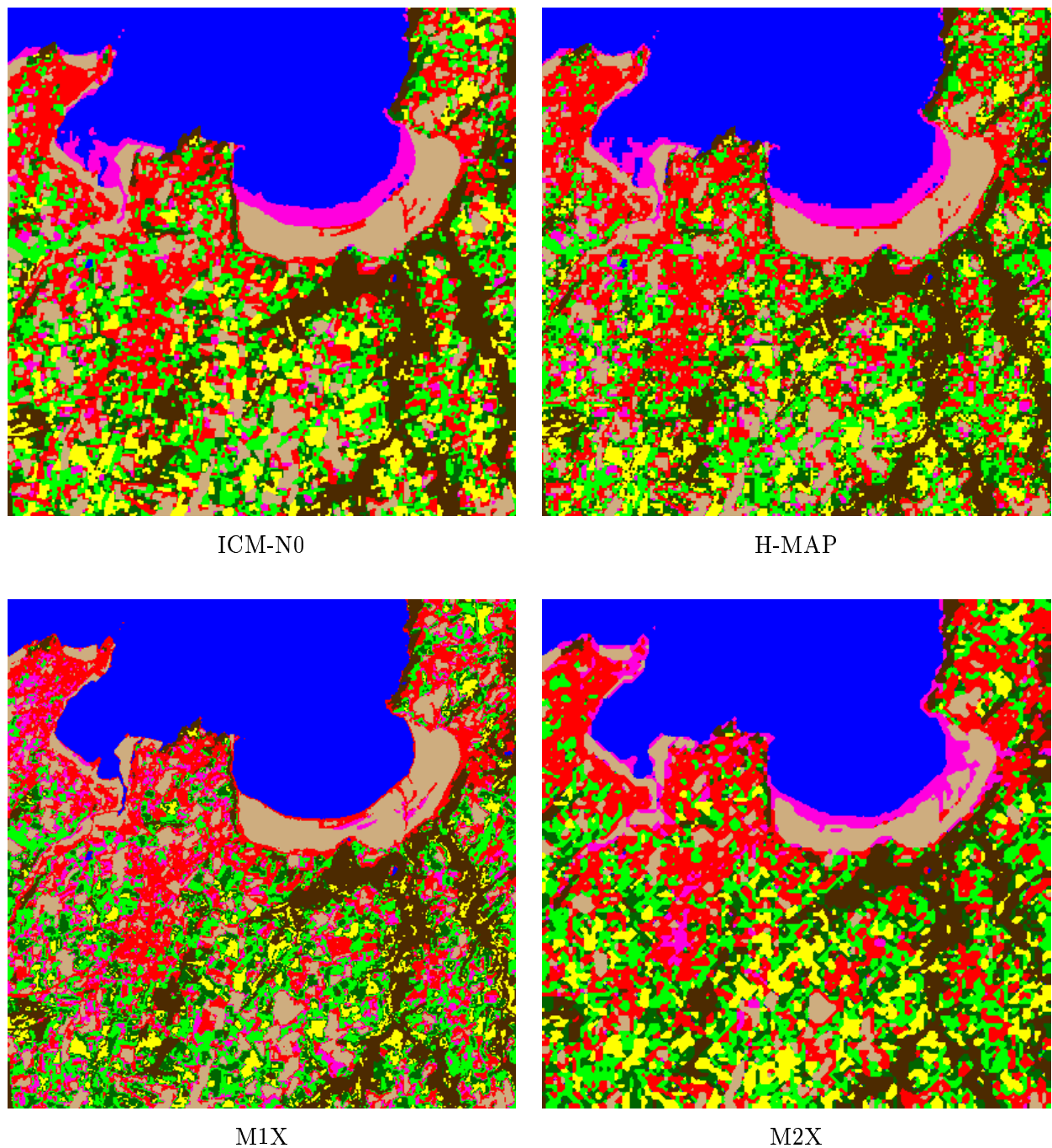


FIG. 6.5 – Résultats de classification pour les modèles ICM-N0, H-MAP, M1X et M2X.

classification (contrainte de niveaux W) est très différent des termes d'attache aux données gaussiens des autres modèles.

Le tableau 6.2 présente les résultats obtenus par les différents modèles en fonction de la vérité terrain. Dans ce tableau, nous donnons le nombre n de pixels mal classés (appelé "erreurs" par la suite) ainsi que le pourcentage p de réussite selon les modèles. Les nombres n et p sont liés par la relation (4.5) donnée dans le chapitre 4. Les modèles MV et ICM-N0 offrent les moins bonnes performances et les trois autres modèles ont des performances du même ordre. Le lecteur intéressé pourra consulter le travail d'A. Chardin [Chardin, 2000] pour une étude comparative approfondie des modèles MV, ICM-N0 et H-MAP.

Nous présentons le détail, classe par classe, des performances des différents modèles dans le tableau 6.3.









Pour la classe "mer et eau", les performances des différents modèles sont bonnes, et très légèrement meilleures pour ce qui est des modèles variationnels M1X et M2X. Le modèle M2X est moins performant que les autres pour la classe "sable et sols nus". Une analyse plus fine montre que cet excès d'erreurs, produit par le modèle M2X par rapport aux autres modèles, provient des petites parcelles de la vérité terrain pour la classe "sable et sols nus" (cf. figure 6.3). Ces parcelles possèdent une taille inférieure à la résolution maximale du modèle M2X, laquelle est de l'ordre de 5×5 pixels. Il en est de même pour la classe "bois et landes" dont la vérité terrain contient (seulement) 14 pixels et est constituée de deux parcelles de tailles respectives 5×1 et 3×3 pixels. Le modèle H-MAP est celui qui offre les meilleures performances en ce qui concerne la classe "zones urbaines". Les villes sont très texturées, et une extraction performante des zones urbaines nécessite de prendre en compte des paramètres de texture comme cela a été effectué dans [Lorette, 1999]. Les différents modèles de classification que nous avons comparés ne prennent pas de tels paramètres de texture en compte, cependant il semble que la structure hiérarchique du modèle H-MAP soit plus robuste que les autres en ce qui concerne la classification des zones urbaines. Les prairies artificielles et naturelles ont un fort chevauchement spectral et sont difficiles à séparer. Il en résulte de mauvaises performances pour ces deux classes, en particulier pour la classe "prairies artificielles" dont les éléments ont tendance à être attribués à la classe "prairie naturelle". Les performances sur la classe "maïs" sont les meilleures pour le modèle M2X qui dépasse les performances des autres modèles d'environ 20% ou plus.

6.3 Conclusion

Nous avons mené une expérience sur des données multispectrales SPOT pour lesquelles nous disposions d'une série de résultats ainsi que d'une vérité terrain. Nous avons comparé

et analysé en détails les résultats obtenus par les différents modèles. Notons que les résultats fournis le projet VISTA ont été obtenus sur des images taille 1480×1024 pixels, alors que nous avons testé nos modèles sur des portions 400×400 pixels afin d'accélérer les calculs. Nous avons évalué les performances des différents modèles en terme de pixels mal classés, grâce à la vérité terrain fournie par les géographes du laboratoire COSTEL. Il est important de souligner le fait que l'avis d'un expert doit faire partie du processus de validation. En effet, d'une part la vérité terrain ne contient que 1466 pixels sur 160000, ce qui est relativement peu pour en tirer des conclusions générales, d'autre part nous pouvons constater que les résultats sont visuellement différents au niveau de la régularité morphologique des régions. La connaissance d'un expert est donc d'une grande importance pour juger de la qualité des résultats et pour déterminer le modèle de classification le mieux adapté pour une application donnée.

CLASSES

	1	2	3	4	5	6	7	8	total
	 mer et eau	 sable et sols nus	 zones urbaines	 bois et landes	 prairies naturelles	 prairies artificielles	 légumes	 maïs	
MV	16 (95.7%)	77 (75.9%)	119 (63.2%)	1 (92.9%)	54 (49.1%)	182 (4.2%)	5 (0%)	58 (56.1%)	512 (65.07%)
ICM-N0	16 (95.7%)	73 (77.2%)	103 (68.1%)	1 (92.9%)	50 (52.8%)	189 (0.5%)	5 (0%)	61 (53.8%)	498 (66.03%)
H-MAP	16 (95.7%)	76 (76.3%)	63 (80.5%)	1 (92.9%)	41 (61.3%)	177 (6.8%)	5 (0%)	54 (59.1%)	433 (70.46%)
M1X	0 (100%)	79 (75.3%)	93 (71.2%)	2 (85.7%)	52 (50.9%)	146 (23.2%)	4 (20%)	44 (66.7%)	421 (71.28%)
M2X	0 (100%)	105 (67.2%)	101 (68.8%)	9 (35.7%)	41 (61.32%)	160 (15.8%)	5 (0%)	19 (85.6%)	440 (70.0%)
nb pixels pour VT	376	320	323	14	106	190	5	132	1466

TAB. 6.3 – Pour chaque classe : nombre de pixels **mal** classés et pourcentage de **réussite** selon les modèles. La dernière ligne comptabilise le nombre de pixels d'échantillons par classe pour la vérité terrain (VT).

Conclusions et perspectives

Principales contributions

Au cours de ce travail, nous avons contribué au développement de modèles variationnels pour la classification d'images. Nous avons proposé deux modèles issus de différentes approches qui conduisent à un étiquetage régularisé des données. Ces modèles ont été étendus au cas de données multispectrales et les résultats obtenus sur une application réelle ont été comparés avec ceux fournis par des approches stochastiques récemment développées.

Le premier modèle variationnel repose sur une suite de minimisation de fonctionnelles dont les solutions convergent vers une partition régulière composée de régions homogènes séparées par des frontières de longueurs minimales. Le cadre théorique, sur lequel est fondée cette approche fait appel à des notions de Γ -convergence, une convergence variationnelle utilisée dans de nombreux problèmes impliquant des ensembles de discontinuités pour approcher un critère contenant un terme lié à la longueur des discontinuités par une suite de fonctionnelles. La suite de critères que nous minimisons contient trois termes : un terme d'attache aux données, un terme de restauration permettant de lisser l'image en préservant les contours ainsi qu'un terme de contrainte de classification tenant compte des caractéristiques des classes. Lors de la convergence de cette suite de critères, le modèle change progressivement de comportement en commençant par restaurer l'image avant d'entamer le processus d'étiquetage des pixels. L'extension de ce modèle au cas de données multispectrales permet de prendre en compte la corrélation entre les différentes bandes à travers un calcul des normes des gradients à l'aide de la géométrie riemannienne, ainsi qu'à travers les covariances des classes prises en compte dans le terme de classification. Nous avons réalisé de nombreuses expérimentations sur un grand nombre de données synthétiques et réelles afin d'étudier les performances de cet algorithme.

Parallèlement à cette approche, nous avons développé un autre modèle variationnel qui consiste à partitionner et étiqueter l'image à l'aide d'un ensemble de régions et contours

dynamiques. Cette approche entre dans le cadre des modèles dynamiques dont les critères comportent des termes reliés à l'information sur les contours de l'image d'une part, mais également sur les régions et leurs caractéristiques d'autre part. Nous utilisons une formulation par ensembles de niveaux permettant de s'affranchir des difficultés liées à un changement de topologie des régions actives et de leurs frontières. Les forces qui guident l'évolution des régions conduisent à une partition des données qui, comme pour le premier modèle, se caractérise par un ensemble de classes homogènes dont la longueur des frontières est minimisée. L'extension au cas multispectral est plus directe que pour le premier modèle proposé, et ne concerne que le terme d'attache aux données. Nous avons délibérément illustré la présentation de ce modèle par un grand nombre de résultats montrant l'évolution des régions et de leurs contours sur divers types de données.

Nous avons présenté une application réelle au sein du dernier chapitre sur des données multispectrales de la baie de Lannion provenant du satellite SPOT 2. Ces résultats ont été étudiés en détail grâce à la vérité terrain mise à notre disposition, et comparés à un modèle stochastique hiérarchique développé récemment au sein du projet VISTA.

Les comparaisons que nous avons menées au sein du chapitre 4 ainsi que dans l'application du chapitre 6 permettent de mettre en avant les propriétés de chacun des deux modèles proposés. Nous avons comparé les deux modèles entre eux, mais également avec des approches stochastiques pour mettre en évidence les avantages et inconvénients de chacun. Nous résumons dans le tableau 6.4 les avantages et inconvénients des deux modèles variationnels proposés. Le modèle de classification, fondé sur une suite de minimisation de fonctionnelles (et faisant appel à des notions de Γ -convergence), est noté M1, alors que le modèle impliquant un ensemble de régions et contours dynamiques est noté M2. Dans le cas de données peu bruitées, nous préférons utiliser le modèle M1 qui a de meilleures performances que le modèle M2 à tous les niveaux (la vitesse de convergence est indépendante du nombre de classes, le lissage préserve les contours, le potentiel W permet de bien gérer un mélange de classes à travers la largeur des puits, les paramètres ont des plages de valeurs acceptables relativement larges), avec quelques précautions à prendre dans le cas multispectral (la vitesse de décroissance de ϵ doit être lente). Lorsque le niveau de bruit est élevé¹, le modèle M2 a l'avantage d'être moins sensible au bruit, grâce à un modèle de contour plus élaboré² que pour M1 et grâce à une initialisation des régions actives dans des

¹"élevé" est pris dans le sens où la valeur des discontinuités dues au bruit est de même ordre que celles liées aux contours des objets.

²"élaboré" car le contour est représenté par une courbe fermée, et non assimilé à de forts gradients des niveaux de gris comme pour le modèle M1, permettant d'être moins sensible aux pixels isolés dus au bruit

	M1 : classif. avec restauration	M2 : classif. par modèle dynamique
nombre de classes K	☺ K puits pour le potentiel W	☺ système de K EDP à résoudre
niveau de bruit	☺ lissage anisotrope	☺ objets de forte courbure
	☺ modèle de contour simple	☺ modèle de contour élaboré
mélange de classes	☺ largeur des puits de W	☺ init. automatique/résolution spatiale
	☺ choix de la décroissance de ε	☺ géré par l'attache aux données
réglage des paramètres		fréq. de régul. des fct. de distance Φ_i
	☺ extension du potentiel W délicate	☺ extension quasi immédiate
	☺ concerne tous les termes de la fonctionnelle	☺ ne concerne que l'attache aux données
cas multispectral	☺ décroissance de ε lente	☺ pas plus de difficulté que dans le cas mono.
temps d'exécution	☺ indépendant du nombre de classes	☺ dépend du nombre de classes

TAB. 6.4 – Avantages et inconvénients des deux modèles de classification développés : M1 est le modèle combinant restauration et classification, M2 est le modèle dynamique de classification.

fenêtres dont la taille est liée à celle des objets qu'il sera possible de détecter et d'étiqueter. De plus, le modèle M2 dans le cas multispectral ne comporte pas plus de difficultés de mise en œuvre que dans le cas monospectral. Cependant, le modèle M2 offre moins de précision que le modèle M1 quant à la détection d'objets de forte courbure, ceci étant dû au fait que dans le cas d'un fort bruit nous devons choisir une forte contrainte sur la longueur de contours dans le modèle M2. Ainsi, dans le cas d'un bruit élevé sur les données, les erreurs commises par M1 sont du type *fausses alarmes* alors que pour le modèle M2 les erreurs proviennent plus particulièrement de cibles non détectées.

Le choix d'un modèle se décide alors en fonction des désirs de l'utilisateur, du type de données et de l'application traitée. Notons au passage que la validation des résultats pour des données réelles est assez délicate dans le sens où elle repose souvent sur des critères subjectifs. Selon les applications, les critères de validation peuvent différer de façon importante. Si l'on prend comme exemple la cartographie, deux experts cartographes réalisant une classification manuelle sur les mêmes données peuvent obtenir deux résultats dont l'étiquetage diffère jusqu'à 10%! Ce qui nous amène à dire que la validation d'un modèle sur des données réelles dépend en quelque sorte de la "vérité" terrain utilisée.

Perspectives

Nous envisageons un certain nombre de perspectives aux travaux présentés dans cette thèse, que ce soit d'ordre pratique ou d'ordre théorique.

Les deux modèles que nous avons proposés entrent dans le cadre de la classification supervisée : le nombre de classes K ainsi que leurs paramètres $(\mu_i, \sigma_i)_{i=1,\dots,K}$ sont supposés connus. Nous avons décrit l'algorithme de *fuzzy Cmeans* modifié nous servant à estimer ces $2K + 1$ paramètres. Afin d'étendre les modèles variationnels proposés au cas de la classification non supervisée, nous introduisons la variable $\Theta = (K, \mu_1, \dots, \mu_K, \sigma_1, \dots, \sigma_K)$ parmi les inconnues des modèles M1 et M2. Ainsi, nous pourrions optimiser alternativement la fonctionnelle J_ϵ (respectivement F_α) du modèle M1 (resp. du modèle M2) par rapport à f (resp. par rapport à $\{\Phi_i\}_{i=1\dots K}$), puis par rapport à Θ . Il est également possible de dissocier le problème de l'estimation du nombre de classes de celui des paramètres des classes. Optimiser les fonctionnelles par rapport aux paramètres des classes en annulant leur gradient peut être raisonnablement envisagé. En revanche, dériver les deux fonctionnelles par rapport à la variable K semble relativement difficile. Pour cela, nous pouvons penser à un algorithme d'optimisation ne nécessitant pas de calcul de dérivées, comme par exemple un algorithme stochastique de type recuit simulé avec un des deux échantillonneurs pré-

sentés au chapitre 1 ou comme l'algorithme du simplex [Nelder et Mead, 1965, Rykov, 1980].

Nous nous sommes intéressés dans ce manuscrit à la classification d'images par rapport aux niveaux de gris des données. Cependant, les modèles de classification que nous avons développés peuvent s'étendre à des critères différents comme des attributs de texture par exemple (utilisation de filtres de Gabor [Jain et Farrokhnia, 1991] ou des matrices de co-occurrence [Conners *et al.*, 1984]), le tout étant que les attributs considérés puissent être caractérisés par des valeurs numériques. Dans le cas des filtres de Gabor, nous pouvons manipuler les réponses des différents filtres, utilisés pour caractériser la texture des données, comme les différentes bandes d'une scène multispectrale. La classification s'effectuera alors sur les réponses des filtres, et non sur les données elles-mêmes.

Les fonctionnelles que nous avons introduites comportent différents termes modélisant les contraintes que l'on souhaite imposer à la solution de classification. Selon le type de données, le niveau de bruit ou encore la régularité de la solution désirée, nous devons ajuster le poids de chaque terme par rapport aux autres. Le réglage de ces paramètres (appelés "hyperparamètres") est parfois délicat et a été opéré manuellement tout au long des expérimentations conduites. Il existe des méthodes pour estimer ces paramètres, faisant souvent appel à des outils de statistique, comme par exemple dans [Hidgon *et al.*, 1997, Jalobeanu *et al.*, 1999, Lakshmanan et Derin, 1989] pour la restauration d'images. L'estimation des hyperparamètres des modèles que nous avons proposés fait partie des axes de recherches envisagés pour la suite de ce travail.

Au niveau algorithmique, nous pourrions améliorer la vitesse de convergence du modèle dynamique M2. Comme nous l'avons précisé dans le chapitre 3, la mise à jour des K EDP engendrées par ce modèle (cf. algorithme 3.2) s'effectue dans une bande étroite autour du niveau zéro des fonctions Φ_i . Cette bande étroite est définie de façon intrinsèque au modèle et non introduite au niveau algorithmique (cf. algorithme de la bande étroite, ou *narrow band* en anglais [Chopp, 1993]), comme c'est souvent le cas pour les modèles actifs formulés par ensembles de niveaux. Cependant, afin d'améliorer la vitesse de convergence, nous pourrions changer de dynamique concernant la mise à jour des EDP dans la bande. Nous pensons par exemple à une approche de type *fast marching* [Sethian, 1999], ou encore à l'algorithme Hermes proposé dans [Paragios et Deriche, 1998] combinant l'approche de type *narrow band* à celle de type *fast marching*.

De plus, nous devons réinitialiser régulièrement les fonctions de distance Φ_i au cours de leur évolution afin de maintenir la contrainte $|\nabla\Phi_i| = 1$. Cette régularisation est effectuée par

l'EDP (3.7). Afin d'éviter cette étape de régularisation, laquelle est coûteuse en temps de calculs, il serait utile d'introduire cette contrainte directement dans le modèle. Dans [Gomes et Faugeras, 1999], des travaux ont été menés pour développer une EDP gouvernant l'évolution de contours actifs par courbure moyenne avec préservation de cette contrainte, ce que ne permet pas l'équation classique de type Hamilton-Jacobi proposée dans [Osher et Sethian, 1988]. Cependant, il en résulte d'autres difficultés qui induisent un schéma algorithmique malgré tout assez coûteux (voir [Gomes et Faugeras, 1999, Gomes et Faugeras, 2000] pour plus de détails).

Au niveau théorique, nous avons soulevé un certain nombre de problèmes au sein des chapitres précédents. Les questions que nous nous posons concernent principalement l'existence (et l'unicité) de solutions aux modèles proposés. Concernant le modèle combinant restauration et classification, nous avons formellement caractérisé les propriétés de la solution du problème limite (expression de la fonctionnelle limite et de sa solution de minimisation) grâce à la théorie de Van der Waals Cahn-Hilliard et à l'aide de la Γ -convergence. Cependant, compte tenu des résultats obtenus (cf. figure 2.7), nous avons remplacé le terme de régularisation quadratique $\varepsilon|\nabla f|^2$ par le terme $\varepsilon^2\varphi(|\nabla f|)$ avec φ une fonction régularisante comme celles définies dans le tableau 2.1. Or, lorsque φ est différente de la fonction quadratique, l'étude des solutions de (2.12) reste un problème ouvert. Une collaboration avec Riccardo March, ainsi que d'autres chercheurs du l'Institut de Mathématiques Appliquées du CNR à Rome, dans le cadre d'une action intégrée Galilée, va permettre d'étudier ce problème théorique.

Par ailleurs, il existe quelques similitudes entre la fonctionnelle J_ε définie dans le cadre de la théorie de Van der Waals Cahn-Hilliard, et les modèles notés Φ^4 définis en physique statistique pour approcher, en quelque sorte, un modèle d'Ising binaire et discret par un modèle continu [Georgii, 1988, Bellac, 1988]. A l'heure actuelle, nous nous sommes contentés de constater ces similitudes, mais il pourrait être intéressant de mener une étude plus approfondie, en collaboration avec des spécialistes du domaine de la physique statistique. L'existence de solutions au modèle de classification par régions et contours actifs reste également un problème ouvert. De plus, nous nous interrogeons sur une extension possible de ce modèle au problème plus général de la segmentation tel qu'il a été proposé par Mumford et Shah [Mumford et Shah, 1985]. Dans ce cas, les régions de la partition cherchée ne possèdent pas de contrainte sur la répartition de l'intensité.

Annexe A

Sur l'équation d'Euler-Lagrange

Le rôle de cette annexe n'est pas de rentrer dans des considérations théoriques concernant les conditions d'optimalité associées aux problèmes de minimisation, nous renvoyons pour cela le lecteur à [Brezis, 1993, Cea, 1971, Giaquinta et Hildebrandt, 1996], mais plutôt de présenter quelques généralités sur l'équation d'Euler-Lagrange, dans le cas monodimensionnel, que nous retrouvons à plusieurs reprises dans cette thèse.

Considérons la fonctionnelle suivante :

$$J(f) = \int_{\Omega} \mathcal{L}(f, f', x) dx, \quad (\text{A.1})$$

où $\Omega \subset \mathbb{R}$ est ouvert borné, $f \in C^1(\Omega, \mathbb{R})$ et $f(x) = g(x), \forall x \in \partial\Omega$.

On pose :

$$f_{\tau}(x) = f(x) + \tau \rho(x), \forall x \in \Omega \quad (\text{A.2})$$

avec $\tau \in \mathbb{R}$, $\rho \in C^1(\Omega, \mathbb{R})$ et

$$\rho(x) = 0, \forall x \in \partial\Omega. \quad (\text{A.3})$$

La fonction f optimise J si :

$$\nabla J(f) = \left. \frac{\partial J}{\partial \tau}(f_{\tau}) \right|_{\tau=0} = 0. \quad (\text{A.4})$$

Or,

$$\nabla J(f) = \int_{\Omega} [\mathcal{L}(f_{\tau}, f'_{\tau}, x) - \mathcal{L}(f, f', x)] dx, \quad (\text{A.5})$$

un développement de \mathcal{L} en série de Taylor, à l'ordre 1 et en négligeant les termes d'ordres supérieurs, donne :

$$\mathcal{L}(f_\tau, f'_\tau, x) - \mathcal{L}(f, f', x) = \tau \frac{\partial \mathcal{L}}{\partial f} \rho + \tau \frac{\partial \mathcal{L}}{\partial f'} \rho'. \quad (\text{A.6})$$

Ainsi, nous avons

$$\nabla J(f) = 0 \Rightarrow \int_{\Omega} \left[\frac{\partial \mathcal{L}}{\partial f} \rho + \frac{\partial \mathcal{L}}{\partial f'} \rho' \right] dx = 0, \quad (\text{A.7})$$

en intégrant par parties, et compte tenu de (A.3), nous obtenons :

$$\int_{\Omega} \frac{\partial \mathcal{L}}{\partial f'} \rho' dx = \left[\frac{\partial \mathcal{L}}{\partial f'} \underbrace{\rho}_{=0} \right]_{\partial \Omega} - \int_{\Omega} \frac{d}{dx} \left(\frac{\partial \mathcal{L}}{\partial f'} \right) \rho dx. \quad (\text{A.8})$$

Par conséquent

$$\nabla J(f) = 0 \Rightarrow \int_{\Omega} \left[\frac{\partial \mathcal{L}}{\partial f} - \frac{d}{dx} \left(\frac{\partial \mathcal{L}}{\partial f'} \right) \right] \rho dx = 0, \quad \forall \rho. \quad (\text{A.9})$$

La relation (A.9) étant vraie pour toute fonction ρ vérifiant (A.3), on en déduit

$$\frac{\partial \mathcal{L}}{\partial f} - \frac{d}{dx} \left(\frac{\partial \mathcal{L}}{\partial f'} \right) = 0. \quad (\text{A.10})$$

L'équation (A.10) est appelée *équation d'Euler-Lagrange* et donne une condition nécessaire pour que f soit solution extrémale de J .

Annexe B

Premier modèle, premiers pas

Nous présentons une première ébauche du modèle proposé dans le chapitre 2. Cette approche intuitive manque d'assise théorique. Néanmoins, les résultats obtenus sont intéressants et méritent d'être mentionnés.

Le modèle proposé au chapitre 2 repose sur la définition d'un critère général de la forme (cf. équation 2.11) :

$$J_\varepsilon(f) = \underbrace{\int_{\Omega} \left(f(x) - I(x)\right)^2 dx}_{\text{attache aux données}} + \varepsilon \lambda^2 \underbrace{\int_{\Omega} \varphi(|\nabla f(x)|) dx}_{\text{terme de restauration}} + \frac{\eta^2}{\varepsilon} \underbrace{\int_{\Omega} W(f(x); \vec{\mu}, \vec{\sigma})}_{\text{contrainte de niveaux}}. \quad (\text{B.1})$$

Avant de nous placer dans le cadre théorique de la théorie de Van der Waals-Cahn-Hilliard pour justifier notre approche, nous avons pensé à définir une fonctionnelle qui diffère de (B.1) par le terme de contrainte de niveaux :

$$J_\varepsilon(f) = \underbrace{\int_{\Omega} \left(f(x) - I(x)\right)^2 dx}_{\text{attache aux données}} + \varepsilon \lambda^2 \underbrace{\int_{\Omega} \varphi_r(|\nabla f(x)|) dx}_{\text{terme de restauration}} + \frac{\eta^2}{\varepsilon} \underbrace{\int_{\Omega} \sum_{i=1}^K \varphi_c\left(\frac{(f(x) - \mu_i)^2}{2\sigma_i^2}\right)}_{\text{contrainte de niveaux}}, \quad (\text{B.2})$$

Avec φ_r et φ_c des fonctions régularisantes ayant le même type de comportement que celui des fonctions régularisantes définies dans le chapitre 2 (cf. tableau 2.1). Ainsi, au lieu de considérer un potentiel de contrainte de niveaux à travers le potentiel W , nous avons défini une contrainte de niveaux sur la fonction f à travers une somme de fonctions régularisantes. Ce potentiel ne retourne pas une valeur nulle pour chaque minimum situé sur les moyennes des classes, et n'est pas localement quadratique autour de chaque minimum comme l'illustre la figure B.1. Par conséquent, ce potentiel ne vérifie pas les conditions requises pour pouvoir se ramener aux travaux effectués dans le cadre de la théorie de Van

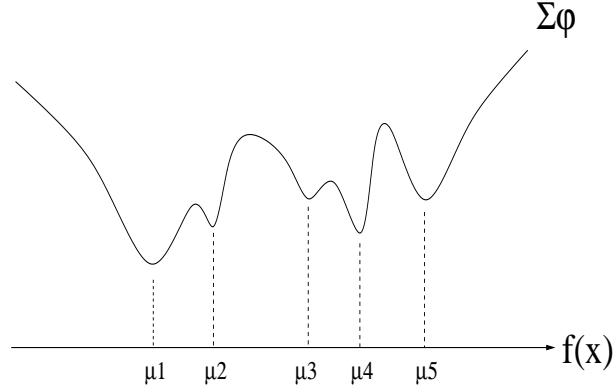


FIG. B.1 – Exemple de potentiel $\sum_{i=1}^K \varphi_c\left(\frac{(f(x)-\mu_i)^2}{2\sigma_i^2}\right)$ dans le cas de 5 classes.

der Waals-Cahn-Hilliard sur les transitions de phases et présentés au sein du paragraphe 2.3. En se référant aux travaux de [Charbonnier *et al.*, 1997, Geman et Reynolds, 1992], et comme pour (2.21) nous avons utilisé la méthode de régularisation semi-quadratique pour minimiser (B.2). On obtient alors une *énergie augmentée* définie sur $(K + 2)$ variables :

$$J_\epsilon^*(f, b, w_1, \dots, w_K) = \int_\Omega (f - I)^2 dx + \epsilon \lambda^2 \int_\Omega [b |\nabla f|^2 + \psi_r(b)] dx \quad (\text{B.3})$$

$$+ \frac{\eta^2}{\epsilon} \sum_{i=1}^K \int_\Omega \left[w_i \left(\frac{(f - \mu_i)^2}{2\sigma_i^2} \right) + \psi_r(w_i) \right] dx$$

Avec ψ_r et ψ_c déduites analytiquement de φ_r et φ_c d'après les travaux de [Charbonnier *et al.*, 1997, Geman et Reynolds, 1992]. Nous minimisons alors J_ϵ^* par rapport à chaque variable en considérant les autres fixées. Ainsi, par le biais des équations d'Euler-Lagrange, nous sommes conduits à résoudre le système suivant pour chaque valeur du paramètre ϵ :

$$\begin{cases} (f - I) + \frac{\eta^2}{\epsilon} \sum_{i=1}^K w_i \frac{(f - \mu_i)}{\sigma_i^2} - \lambda^2 \epsilon \operatorname{div}(b \nabla f) = 0 & \text{pour } b, w_1, \dots, w_K \text{ fixés} \\ b = \frac{\varphi'_c(|\nabla f|)}{2|\nabla f|} & \text{pour } f \text{ fixé} \\ w_i = \frac{\varphi'_c\left(\frac{(f - \mu_i)^2}{2\sigma_i^2}\right)}{2\frac{(f - \mu_i)^2}{2\sigma_i^2}} & \text{pour } f \text{ fixé, et pour } i = 1..K \end{cases} \quad (\text{B.4})$$

La variable b est liée à la présence de contours ($b(x) \simeq 0$ si x est sur un contour et $b(x) \simeq 1$ sur une zone homogène), alors que les K variables w_i représentent les degrés d'appartenance

aux classes ($w_i(x) \simeq 1$ si x est proche de μ_i et $w_i(x) \simeq 0$ si la valeur x est loin de celle de μ_i).

Ce modèle n'a pas de justification théorique mais les résultats obtenus sont sensiblement comparables à ceux obtenus en utilisant un potentiel de contrainte de niveaux W pour les images considérées. Il semble évident que le potentiel défini à travers la somme de fonctions régularisantes dans (B.2) favorise les classes du fait de la différence de profondeur des puits (cf. illustration de la figure B.1). Les pixels qui tombent dans les puits les plus profonds minimisent plus fortement l'énergie que ceux situés dans les puits les moins profonds. Ce problème peut être résolu par une renormalisation du potentiel afin d'avoir des puits de profondeurs identiques.

Nous ne nous étendrons pas plus sur ce modèle car nous ne l'avons pas exploité plus loin même s'il mériterait une étude plus approfondie. Nous présentons quand même deux résultats obtenus sur une image synthétique (cf. figure B.2) ainsi que sur une image SPOT, canal XS3 (cf. figure B.3).

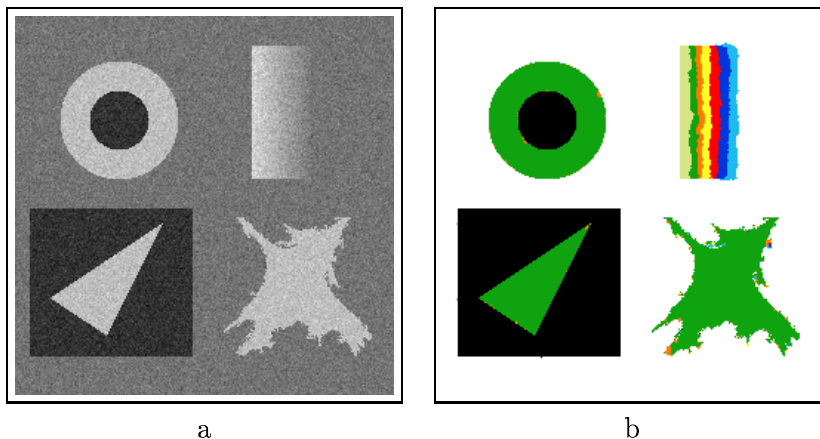


FIG. B.2 – Résultats de classification en fausses couleurs dans le cas de 9 classes (figure b) sur des données synthétiques bruitées à 10dB fournies par le GDR-PRC ISIS (figure a).

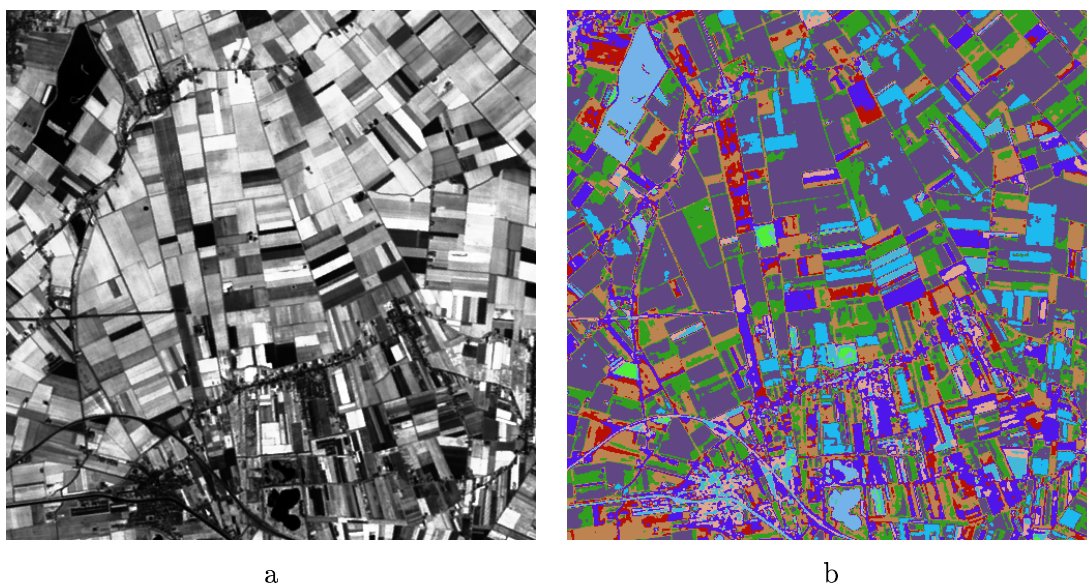


FIG. B.3 – Résultats de classification en fausses couleurs dans le cas de 10 classes (figure b) sur des données SPOT fournies par le CNES (figure a).

Annexe C

Discrétisation du premier modèle

Cette annexe est relative à la discrétisation des équations présentées dans le schéma algorithmique du tableau 2.2.

Pour chaque valeur de ϵ_l , nous résolvons le système suivant :

$$S_{\epsilon_l}(b, f) \begin{cases} b = \frac{\varphi'(|\nabla f|)}{2|\nabla f|} \\ f - \lambda^2 \epsilon_l \operatorname{div}(b \nabla f) + \frac{\eta^2}{2\epsilon_l} W'(f) = I \end{cases} \quad (\text{C.1})$$

qui admet pour solution le couple $(b_{\epsilon_l}, f_{\epsilon_l})$. Nous obtenons une famille de solutions $(b_{\epsilon_l}, f_{\epsilon_l})$ à la suite de systèmes $\{S_{\epsilon_l}\}_{\epsilon_l}$ lorsque ϵ_l diminue selon la loi $\epsilon_l = \rho \epsilon_{l-1}$ avec $\rho < 1$. Chaque système $S_{\epsilon_{l+1}}$ est initialisé par le couple $(b_{\epsilon_l}, f_{\epsilon_l})$, solution de S_{ϵ_l} à l'étape précédente.

Nous allons donner l'expression discrète du système d'équations S_{ϵ_l} en un point (m, n) de la grille de pixels. Les discrétisations adoptées sont du même type que celles utilisées dans [Charbonnier, 1994, Teboul, 1999].

- Expression discrète de b :

Nous utilisons une discrétisation décentrée pour le calcul de $|\nabla f|$, si bien que :

$$b_{mn} = \frac{\varphi' \left(\sqrt{(f_{mn+1} - f_{mn})^2 - (f_{m+1n} - f_{mn})^2} \right)}{2\sqrt{(f_{mn+1} - f_{mn})^2 - (f_{m+1n} - f_{mn})^2}}.$$

- Expression discrète de $W'(f(x))$:

Le potentiel W est parabolique par morceaux comme l'indique l'expression (2.13). Par conséquent W' est linéaire par morceaux (cf. (2.22)), et pour chaque valeur f_{mn} , il existe

deux réels A_{mn} et B_{mn} dont les valeurs sont fonctions de celles de f_{mn} :

$$W'(f_{mn}) = 2[A_{mn}f_{mn} + B_{mn}].$$

- Expression discrète de $\text{div}(b\nabla f)$:

La discrétisation adoptée pour ∇f est :

$$\nabla f_{mn} = (f_{mn+1} - f_{mn}, f_{m+1n} - f_{mn}).$$

Ainsi,

$$\begin{aligned} \text{div}(b\nabla f)_{mn} &= \text{div}(b_{mn}(f_{mn+1} - f_{mn}), b_{mn}(f_{m+1n} - f_{mn})) \\ &= b_{mn}(f_{mn+1} - f_{mn}) - b_{mn-1}(f_{mn-1} - f_{mn}) \\ &\quad + b_{mn}(f_{m+1n} - f_{mn}) - b_{m-1n}(f_{m-1n} - f_{mn}). \end{aligned}$$

Par conséquent, l'expression discrète du système S_{ϵ_l} (C.1) au pixel de coordonnées (m, n) s'écrit :

$$S_{\epsilon_l}(b_{mn}, f_{mn}) \begin{cases} b_{mn} = \frac{\varphi' \left(\sqrt{(f_{mn+1} - f_{mn})^2 - (f_{m+1n} - f_{mn})^2} \right)}{2\sqrt{(f_{mn+1} - f_{mn})^2 - (f_{m+1n} - f_{mn})^2}} \\ I_{mn} = f_{mn} - \lambda^2 \epsilon_l \left\{ b_{mn}(f_{mn+1} - f_{mn}) - b_{mn-1}(f_{mn-1} - f_{mn}) \right. \\ \quad \left. + b_{mn}(f_{m+1n} - f_{mn}) - b_{m-1n}(f_{m-1n} - f_{mn}) \right\} + \frac{\eta^2}{\epsilon_l} \{ A_{mn}f_{mn} + B_{mn} \} \end{cases}$$

Annexe D

Démonstration du lemme 1

Lemme :

Soit

$$\mathcal{L}_\alpha(\Phi_i) = \int_{\Omega} g(|\nabla I|) \delta_\alpha(\Phi_i(x)) |\nabla \Phi_i(x)| dx,$$

alors nous avons

$$\lim_{\alpha \rightarrow 0} \mathcal{L}_\alpha(\Phi_i) = \int_{\Phi_i=0} g(|\nabla I|) ds,$$

Preuve :

D'après la formule de la co-aire, [Evans et Gariepy, 1992], nous savons que

$$\mathcal{L}_\alpha(\Phi_i) = \int_{\mathbb{R}} \left[\delta_\alpha(\rho) \int_{\Phi_i=\rho} g(|\nabla I|) ds \right] d\rho$$

En posant $h(\rho) = \int_{\Phi_i=\rho} g(|\nabla I|) ds$, nous obtenons

$$\mathcal{L}_\alpha(\Phi_i) = \int_{\mathbb{R}} \delta_\alpha(\rho) h(\rho) d\rho = \frac{1}{2\alpha} \int_{-\alpha}^{\alpha} \left(1 + \cos\left(\frac{\pi\rho}{\alpha}\right) \right) h(\rho) d\rho$$

Si on effectue le changement de variable $\theta = \frac{\rho}{\alpha}$ alors

$$\mathcal{L}_\alpha(\Phi_i) = \frac{1}{2} \int_{-1}^1 \left(1 + \cos(\pi\theta) \right) h(\alpha\theta) d\theta$$

Ainsi, lorsque $\alpha \rightarrow 0$ on a

$$\lim_{\alpha \rightarrow 0} \mathcal{L}_\alpha(\Phi_i) = \frac{1}{2} h(0) \int_{-1}^1 \left(1 + \cos(\pi\theta) \right) d\theta = h(0) = \int_{\Phi_i=0} g(|\nabla I|) ds$$

Annexe E

Calcul des équations (3.36)

Nous allons calculer les équations d'Euler-Lagrange relatives à la minimisation de :

$$F_\alpha(\Phi_1, \dots, \Phi_K) = \sum_{i=1}^K e_i \int_{\Omega} H_\alpha(\Phi_i) \frac{(I(x) - \mu_i)^2}{\sigma_i^2} dx + \sum_{i=1}^K \gamma_i \int_{\Omega} g(|\nabla I|) \delta_\alpha(\Phi_i) |\nabla \Phi_i| dx \\ + \frac{\lambda}{2} \int_{\Omega} \left(\sum_{i=1}^K H_\alpha(\Phi_i) - 1 \right)^2 dx$$

A partir de (3.23), (3.24) et (3.29) nous pouvons écrire

$$\begin{aligned} F_\alpha(\Phi_1, \dots, \Phi_K) &= F_\alpha^A(\Phi_1, \dots, \Phi_K) + F_\alpha^B(\Phi_1, \dots, \Phi_K) + F_\alpha^C(\Phi_1, \dots, \Phi_K) \\ &= \int_{\Omega} \mathcal{L}_\alpha^a(\Phi_1, \dots, \Phi_K, \nabla \Phi_1, \dots, \nabla \Phi_K, x) dx \\ &+ \int_{\Omega} \mathcal{L}_\alpha^b(\Phi_1, \dots, \Phi_K, \nabla \Phi_1, \dots, \nabla \Phi_K, x) dx \\ &+ \int_{\Omega} \mathcal{L}_\alpha^c(\Phi_1, \dots, \Phi_K, \nabla \Phi_1, \dots, \nabla \Phi_K, x) dx \end{aligned}$$

Si (Φ_1, \dots, Φ_K) est solution de la minimisation de F_α , alors nécessairement

$$\frac{\partial F_\alpha}{\partial \Phi_i} = 0, \quad \forall i = 1 \dots K,$$

et les équations d'Euler-Lagrange pour tout $i = 1 \dots K$ (avec des conditions de Neumann aux bords $\partial\Omega$) conduisent à :

$$\begin{aligned} \frac{\partial F_\alpha^c}{\partial \Phi_i} &= \frac{\partial \mathcal{L}_\alpha^c}{\partial \Phi_i} - \operatorname{div} \left(\frac{\partial \mathcal{L}_\alpha^c}{\partial \nabla \Phi_i} \right) \\ &= \lambda \delta_\alpha(\Phi_i) \left(\sum_{j=1}^K H_\alpha(\Phi_j) - 1 \right) \end{aligned}$$

$$\begin{aligned}
\frac{\partial F_\alpha^A}{\partial \Phi_i} &= \frac{\partial \mathcal{L}_\alpha^a}{\partial \Phi_i} - \operatorname{div} \left(\frac{\partial \mathcal{L}_\alpha^a}{\partial \nabla \Phi_i} \right) \\
&= \frac{\partial H_\alpha}{\partial \Phi_i}(\Phi_i) e_i \frac{(u_0 - \mu_i)^2}{\sigma_i^2} \\
&= \delta_\alpha(\Phi_i) e_i \frac{(u_0 - \mu_i)^2}{\sigma_i^2}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial F_\alpha^B}{\partial \Phi_i} &= \frac{\partial \mathcal{L}_\alpha^b}{\partial \Phi_i} - \operatorname{div} \left(\frac{\partial \mathcal{L}_\alpha^b}{\partial \nabla \Phi_i} \right) \\
&= \gamma_i \left[g(|\nabla I|) \delta'_\alpha(\Phi_i) |\nabla \Phi_i| - \operatorname{div} \left(g(|\nabla I|) \delta_\alpha(\Phi_i) \frac{\nabla \Phi_i}{|\nabla \Phi_i|} \right) \right]
\end{aligned}$$

et si $x = (x_1, x_2) \in \Omega$ nous avons

$$\begin{aligned}
\operatorname{div} \left(g(|\nabla I|) \delta_\alpha(\Phi_i) \frac{\nabla \Phi_i}{|\nabla \Phi_i|} \right) &= \frac{\partial}{\partial x_1} \left(g(|\nabla I|) \delta_\alpha(\Phi_i) \frac{\partial \Phi_i}{|\nabla \Phi_i| \partial x_1} \right) + \frac{\partial}{\partial x_2} \left(g(|\nabla I|) \delta_\alpha(\Phi_i) \frac{\partial \Phi_i}{|\nabla \Phi_i| \partial x_2} \right) \\
&= \delta_\alpha(\Phi_i) \frac{\nabla g \nabla \Phi_i}{|\nabla \Phi_i|} \\
&\quad + g(|\nabla I|) \left[\frac{\partial \delta_\alpha}{\partial \Phi_i}(\Phi_i) \frac{\partial \Phi_i}{\partial x_1} \frac{\nabla \Phi_i}{|\nabla \Phi_i|} + \frac{\partial \delta_\alpha}{\partial \Phi_i}(\Phi_i) \frac{\partial \Phi_i}{\partial x_2} \frac{\nabla \Phi_i}{|\nabla \Phi_i|} + \delta_\alpha(\Phi_i) \operatorname{div} \left(\frac{\nabla \Phi_i}{|\nabla \Phi_i|} \right) \right] \\
&= \delta_\alpha(\Phi_i) \frac{\nabla g \nabla \Phi_i}{|\nabla \Phi_i|} + g(|\nabla I|) \left[\delta'_\alpha(\Phi_i) |\nabla \Phi_i| + \delta_\alpha(\Phi_i) \operatorname{div} \left(\frac{\nabla \Phi_i}{|\nabla \Phi_i|} \right) \right]
\end{aligned}$$

donc

$$\frac{\partial F_\alpha^B}{\partial \Phi_i} = -\gamma_i \delta_\alpha(\Phi_i) \left[\frac{\nabla g \nabla \Phi_i}{|\nabla \Phi_i|} + g(|\nabla I|) \operatorname{div} \left(\frac{\nabla \Phi_i}{|\nabla \Phi_i|} \right) \right]$$

Annexe F

Discrétisation du second modèle

Cette annexe est relative à la discrétisation des équations présentées dans le schéma algorithmique du tableau 3.2, en un pixel (m, n) de la grille.

Nous utilisons des discrétisations de même nature que celles données dans [Osher et Sethian, 1988, Zhao *et al.*, 1996]. Les travaux [Krissian, 2000, Sethian, 1996] offrent un bon aperçu des différents schémas algorithmiques utilisés dans l'implantation des méthodes par ensembles de niveaux.

Afin d'alléger les notations, notons $u = \Phi_i$, la fonction d'ensemble de niveaux associée à la classe C_i .

-1) Discrétisation de :

$$\begin{aligned} \frac{\partial u}{\partial t} = - \left\{ \delta_\alpha(u) \left[e_i \frac{(I - \mu_i)^2}{\sigma_i^2} - \gamma_i g(\nabla I) \operatorname{div} \left(\frac{\nabla u}{|\nabla u|} \right) - \gamma_i \frac{\nabla g \nabla u}{|\nabla u|} \right. \right. \\ \left. \left. + \lambda \left(\sum_{j=1}^K H_\alpha(\Phi_j) - 1 \right) \right] \right\}. \end{aligned} \quad (\text{F.1})$$

- i) discrétisation de $g(I)$:

D'après la définition de g nous optons pour la discrétisation suivante :

$$g_{mn} = \frac{1}{1 + (I_{mn+1} - I_{mn})^2 + (I_{m+1n} - I_{mn})^2}$$

- ii) discrétisation de $\frac{\nabla g \nabla u}{|\nabla u|}$:

$$\left(\frac{\nabla g \nabla u}{|\nabla u|} \right)_{mn} = \frac{(g_{mn+1} - g_{mn})(u_{mn+1} - u_{mn}) + (g_{m+1n} - g_{mn})(u_{m+1n} - u_{mn})}{\sqrt{(u_{mn+1} - u_{mn})^2 + (u_{m+1n} - u_{mn})^2}}$$

- *iii*) discrétisation de $div\left(\frac{\nabla u}{|\nabla u|}\right)$:

Suivant [Zhao *et al.*, 1996], nous posons :

$$div\left(\frac{\nabla u}{|\nabla u|}\right)_{mn} = \left(\frac{u_x}{|\nabla u|}\right)_{m+\frac{1}{2}n} - \left(\frac{u_x}{|\nabla u|}\right)_{m-\frac{1}{2}n} + \left(\frac{u_y}{|\nabla u|}\right)_{mn+\frac{1}{2}} - \left(\frac{u_y}{|\nabla u|}\right)_{mn-\frac{1}{2}},$$

avec :

$$\begin{aligned} \left(\frac{u_x}{|\nabla u|}\right)_{m+\frac{1}{2}n} &= \frac{u_{m+1n} - u_{mn}}{\sqrt{(u_{m+1n} - u_{mn})^2 + 0.25[(u_{mn+1} - u_{mn-1})/2 + (u_{m+1n+1} - u_{m+1n-1})/2]^2}} \\ \left(\frac{u_x}{|\nabla u|}\right)_{m-\frac{1}{2}n} &= \frac{u_{mn} - u_{m-1n}}{\sqrt{(u_{mn} - u_{m-1n})^2 + 0.25[(u_{m-1n+1} - u_{m-1n-1})/2 + (u_{mn+1} - u_{mn-1})/2]^2}} \\ \left(\frac{u_y}{|\nabla u|}\right)_{mn+\frac{1}{2}} &= \frac{u_{mn+1} - u_{mn}}{\sqrt{(u_{m+1n} - u_{m-1n})^2 + 0.25[(u_{m+1n+1} - u_{m-1n+1})/2 + (u_{mn+1} - u_{mn})/2]^2}} \\ \left(\frac{u_y}{|\nabla u|}\right)_{mn-\frac{1}{2}} &= \frac{u_{mn} - u_{mn-1}}{\sqrt{(u_{m+1n-1} - u_{m-1n-1})^2 + 0.25[(u_{m+1n} - u_{m-1n})/2 + (u_{mn} - u_{mn-1})/2]^2}} \end{aligned}$$

Ainsi, nous obtenons le schéma discret pour (F.1) au point (m, n) :

$$\begin{aligned} u_{mn}^{t+1} = u_{mn}^t - dt \left\{ \delta_\alpha(u_{mn}^t) \left[e_i \frac{(I_{mn} - \mu_i)^2}{\sigma_i^2} - \gamma_i g_{mn} div\left(\frac{\nabla u^t}{|\nabla u^t|}\right)_{mn} - \gamma_i \left(\frac{\nabla g \nabla u^t}{|\nabla u^t|}\right)_{mn} \right. \right. \\ \left. \left. + \lambda \left(\sum_{j=1}^K H_\alpha(\Phi_{j,mn}^t) - 1 \right) \right] \right\}. \quad (F.2) \end{aligned}$$

- **2**) Discrétisation de l'EDP de régularisation des fonctions u :

Nous devons réinitialiser les fonctions de distance u au cours de l'évolution du programme (cf. tableau 3.2) par l'EDP proposée dans [Sussman *et al.*, 1994] :

$$\frac{\partial u}{\partial t} = signe(u)(1 - |\nabla u|)$$

où

$$signe(u) = \begin{cases} +1 & \text{si } u > 0 \\ -1 & \text{si } u < 0 \\ 0 & \text{si } u = 0 \end{cases}$$

La fonction $signe$ est discrétisée comme suit :

$$signe(u)_{mn} = \frac{u_{mn}}{\sqrt{u_{mn}^2 + \tau}} \quad (F.3)$$

avec $\tau \approx 10^{-6}$ permettant d'éviter une division par zéro.

Posons :

$$\begin{aligned} a &= u_{mn} - u_{m-1n} \\ b &= u_{m+1n} - u_{mn} \\ c &= u_{mn} - u_{mn-1} \\ d &= u_{mn+1} - u_{mn} \end{aligned} \tag{F.4}$$

et nous définissons

$$G(u)_{mn} = \begin{cases} \sqrt{\max((a^+)^2, (b^-)^2) + \max((c^+)^2, (d^-)^2)} - 1 & \text{si } u_{mn} > 0 \\ \sqrt{\max((a^-)^2, (b^+)^2) + \max((c^-)^2, (d^+)^2)} - 1 & \text{si } u_{mn} < 0 \\ 0 & \text{sinon} \end{cases}$$

avec l'exposant $+$ (resp. $-$) désignant la partie positive : $a^+ = \max(a, 0)$ (resp. négative : $a^- = \min(a, 0)$).

Nous obtenons ainsi le schéma discret suivant pour (F.3) :

$$u_{mn}^{t+1} = u_{mn}^t - dt \left\{ \text{signe}(u)_{mn} G(u)_{mn} \right\}. \tag{F.5}$$

Bibliographie

- [Allen et Cahn, 1979] S. Allen et J. Cahn. A microscopic theory for antiphase boundary motion and its application to antiphase domain coarsening. *Acta Metallurgica*, 27 :1085–1095, 1979.
- [Alvarez *et al.*, 1992] L. Alvarez, P.-L. Lions et J.-M. Morel. Image selective smoothing and edge detection by nonlinear diffusion. *SIAM J. of Numerical Analysis*, 29(3) :845–866, 1992.
- [Ambrosio et Tortorelli, 1990] L. Ambrosio et V. Tortorelli. Approximation of functionals depending on jumps by elliptic functionals via Γ -convergence. *Commun. Pure Appl. Math.*, 43 :999–1036, 1990.
- [Ambrosio et Tortorelli, 1992] L. Ambrosio et V. Tortorelli. On the approximation of functionals depending on jumps by quadratic, elliptic functionals. *Boll. Un. Mat. Ital.*, 46-B :105–123, 1992.
- [Angenent et Gurtin, 1989] S. Angenent et M.E. Gurtin. Multiphase thermomechanics with interfacial structure 2. Evolution of an isothermal interface. *Arch. Rational Mech. Anal.*, 108 :333–391, 1989.
- [Aubert et Vese, 1997] G. Aubert et L. Vese. A variational method in image recovery. *SIAM J. of Numerical Analysis*, 34(5) :1948–1979, 1997.
- [Aujol, 1999] J.-F. Aujol. Traitement d’images par contours actifs (snakes) - Conservation de la fonction de distance. Rapport interne, projet ARIANA, INRIA Sophia Antipolis, France, 1999.
- [Baldo, 1990] S. Baldo. Minimal interface criterion for phase transitions in mixtures of Cahn-Hilliard fluids. *Ann. Inst. Henri Poincaré*, 7 :67–90, 1990.
- [Barles *et al.*, 1992] G. Barles, L. Bronsard et P.E. Souganidis. Front propagation for reaction-diffusion equations of bistable type. *Ann. Inst. Henri Poincaré*, 9 :479–496, 1992.
- [Barles *et al.*, 1993] G. Barles, H. M. Soner et P.E. Souganidis. Front propagation and phase field theory. *SIAM J. Control and Optimization*, 31 :439–479, 1993.

- [Basseville, 1985] M. Basseville. Distance measures for signal processing and pattern recognition. *Signal processing*, 18 :349–369, 1985.
- [Baudoin *et al.*, 1996] A. Baudoin, C. Fratter et P. Munier. Le programme SPOT 5 et la préparation des utilisations. *Bul. S.F.P.T.*, 141 :9–21, janvier 1996.
- [Bellac, 1988] M. Le Bellac. *Des phénomènes critiques aux champs de jauge*. InterEditions/Editions du CNRS, 1988.
- [Bellettini *et al.*, 1991] G. Bellettini, M. Paolini et C. Verdi. Numerical minimization of geometrical type problems related to calculus of variations. *Calcolo*, 27 :251–278, 1991.
- [Berthod *et al.*, 1996] M. Berthod, Z. Kato, S. Yu et J. Zerubia. Bayesian image classification using Markov random fields. *Image and Vision Computing*, 14(4) :285–293, 1996.
- [Besag, 1974] J. E. Besag. Spatial interaction and the statistical analysis of lattice system. *Journal of the royal statistical society, Series B*, 36 :192–236, 1974.
- [Besag, 1986] J. E. Besag. On the statistical analysis of dirty pictures. *Journal of the royal statistical society, Series B*, 68 :259–302, 1986.
- [Bezdek, 1981] J.C. Bezdek. *Pattern recognition with fuzzy objective function algorithms*. Plenum Press, New-York, 1981.
- [Bhattacharyya, 1967] C. Bhattacharyya. A simple method of resolution of a distribution into Gaussian components. *Biometrics*, 23(4) :115–135, 1967.
- [Blake et Zisserman, 1987] A. Blake et A. Zisserman. *Visual reconstruction*. M.I.T. Press, 1987.
- [Blomgren et Chan, 1998] P. Blomgren et T.F. Chan. Color TV : Total variation methods for restoration of vector-valued images. *IEEE Trans. on Image Processing*, 7(3) :304–309, Mars 1998. Special issue on partial differential equations and geometry driven diffusion in image processing and analysis.
- [Blomgren, 1998] P. Blomgren. *Color TV : Total Variation Methods for Restoration of Vector-Valued Images*. Thèse de Doctorat, University of California, Los Angeles, 1998.
- [Bouman et Shapiro, 1994] C.A. Bouman et M. Shapiro. A multiscale random field model for Bayesian image segmentation. *IEEE Trans. on Image Processing*, 3 :162–177, Mars 1994.
- [Braides, 1998] A. Braides. *Approximation of free-discontinuity problems*. Lecture notes in mathematics, number 1664, Springer, 1998.
- [Brezis, 1993] H. Brezis. *Analyse fonctionnelle : théorie et applications*. Masson, 1993.
- [Bronsard et Reitich, 1993] L. Bronsard et F. Reitich. On three-phase boundary motion and the singular limit of a vector-valued Ginzburg-Landau equation. *Arch. Rational Mech. Anal.*, 124 :355–379, 1993.

-
- [Caginalp, 1989] G. Caginalp. Stefan and Hele-Shaw type as asymptotic limits of the phase-field equations. *Physical Review*, 39(11) :5887–5896, 1989.
- [Caselles *et al.*, 1993] V. Caselles, F. Catte, T. Coll et F. Dibos. A geometric model for active contours. *Numerische Mathematik*, 66 :1–31, 1993.
- [Caselles *et al.*, 1997] V. Caselles, R. Kimmel et G. Sapiro. Geodesic active contours. *International J. of Computer Vision*, 22(1) :61–79, 1997.
- [Castleman, 1996] K. Castleman. *Digital image processing*. Prentice-Hall, 1996.
- [Catté *et al.*, 1992] F. Catté, P.-L. Lions, J.-M. Morel et T. Coll. Image selective smoothing and edge detection by nonlinear diffusion. *SIAM J. of Numerical Analysis*, 29(1) :182–193, 1992.
- [Cea, 1971] J. Cea. *Optimisation : théorie et algorithmes*. Dunod, 1971.
- [Chalmond, 2000] B. Chalmond. *Éléments de modélisation pour l'analyse d'images*. Springer-Verlag, 2000.
- [Chambolle et Lions, 1997] A. Chambolle et P.L. Lions. Image recovery via total variation minimization and related problems. *Numer. Math.*, 76(2) :167–188, 1997.
- [Chan et Vese, 1999] T. Chan et L. Vese. An active contour model without edges. In *Proc. of IEEE ICCV, Corfu, Grèce*, pages 141–151, Septembre 1999.
- [Chan et Wong, 1998] T. Chan et C. Wong. Total variation blind deconvolution. *IEEE Trans. on Image Processing*, 7(3) :370–375, Mars 1998.
- [Charbonnier *et al.*, 1997] P. Charbonnier, L. Blanc-Féraud, G. Aubert et M. Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Trans. on Image Processing*, 6(2) :298–311, février 1997.
- [Charbonnier, 1994] P. Charbonnier. *Restauration d'image : régularisation avec prise en compte des discontinuités*. Thèse de Doctorat, Université de Nice Sophia Antipolis, France, Septembre 1994.
- [Chardin, 2000] A. Chardin. *Modèles énergétiques hiérarchiques pour la résolution des problèmes inverses en analyse d'images*. Thèse de Doctorat, Université de Rennes I, France, 2000.
- [Chopp, 1993] D. Chopp. Computing minimal surfaces via level set curvature flow. *J. of Computational Physics*, 106 :77–91, 1993.
- [CNES, 1989] CNES. Le catalogue des produits spot et des services. CNES, SPOT IMAGE, 1989.
- [Cocquerez et Philipp, 1995] J.-P. Cocquerez et S. Philipp. *Analyse d'images : filtrage et segmentation*. Masson, Paris, 1995.

- [Cohen, 1991] L. D. Cohen. On active contour models and balloon. *Computer Vision, Graphics, and Image Processing : Image Understanding (CVGIP :IU)*, 53(2) :211–218, 1991.
- [Connors *et al.*, 1984] R. Connors, M. Trivedi et C. Harlow. Segmentation of high-resolution urban scene using texture operators. *Computer Vision, Graphics, and Image Processing (CVGIP)*, 25 :273–310, 1984.
- [Deriche et Faugeras, 1996] R. Deriche et O. Faugeras. Les EDP en traitement des images et vision par ordinateur. *Traitement du signal*, 13(6), 1996.
- [Descombes *et al.*, 1995] X. Descombes, J.-F. Mangin, E. Pechersky et M. Sigelle. Fine structures preserving model for image processing. In *Proc. 9th SCIA Uppsala, Suède*, pages 349–356, 1995.
- [Descombes *et al.*, 1997a] X. Descombes, R. Morris et J. Zerubia. Quelques améliorations à la segmentation d’images bayésiennes - Deuxième partie : classification. *Traitement du Signal*, 14(4) :383–395, 1997.
- [Descombes *et al.*, 1997b] X. Descombes, R. Morris et J. Zerubia. Quelques améliorations à la segmentation d’images bayésiennes - Première partie : modélisation. *Traitement du Signal*, 14(4) :373–382, 1997.
- [Descombes *et al.*, 1999] X. Descombes, R. Morris, J. Zerubia et M. Berthod. Estimation of Markov random field prior parameters using Markov chain Monte Carlo maximum likelihood. *IEEE Trans. on Image Processing*, 8(7) :954–963, 1999.
- [Descombes, 1993] X. Descombes. *Champs markoviens en analyse d’images*. Thèse de Doctorat, Ecole Nationale Supérieure des Télécommunications, France, 1993.
- [Durand *et al.*, 1999] S. Durand, F. Malgouyres et B. Rougé. Image de-blurring, spectrum interpolation and application to satellite imaging. Rapport de Recherche CMLA 9916, (<http://www.cmla.ens-cachan.fr/Cmla/Publications/>), 1999.
- [Evans et Gariepy, 1992] L. C. Evans et R. F. Gariepy. *Measure theory and fine properties of functions*. CRC Press, 1992.
- [Evans et Spruck, 1992] L.C. Evans et J. Spruck. Motion of level sets by mean curvature. part II. *Trans. of the American Mathematical Society*, 330(1) :321–332, 1992.
- [Fonseca et Tartar, 1989] I. Fonseca et L. Tartar. The gradient theory of phase transitions for systems with two potential wells. *Proc. of the Royal Society of Edinburgh*, 111A(11) :89–102, 1989.
- [Frigui et Krishnapuram, 1996] H. Frigui et R. Krishnapuram. A robust clustering algorithm based on competitive agglomeration and soft rejection of outliers. In *CVPR*, pages 550–555, San Francisco, Etats-Unis, juin 1996.

-
- [Fukunaga, 1972] K. Fukunaga. *Introduction to statistical pattern recognition*. Academic Press, New York, 1972.
- [Gadat, 1999] S. Gadat. Restauration d'images satellitaires multispectrales. Rapport interne, projet ARIANA, INRIA Sophia Antipolis, France, 1999.
- [Geman et Geman, 1984] S. Geman et D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 6(6) :721–741, novembre 1984.
- [Geman et McClure, 1985] S. Geman et D.-E. McClure. Bayesian image analysis : an application to single photon emission tomography. *Proc. Statist. Comput. Sect.*, Washington DC, Amer. Statist. Assoc. :11–18, 1985.
- [Geman et Reynolds, 1992] D. Geman et G. Reynolds. Constrained restoration and the recovery of discontinuities. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(3) :367–383, 1992.
- [Georgii, 1988] H. Georgii. *Gibbs measures and phase transitions*. De Gruyter - studies in mathematics, vol. 9, 1988.
- [Giaquinta et Hildebrandt, 1996] M. Giaquinta et S. Hildebrandt. *Calculus of variations (vol. 1,2)*. Springer, 1996.
- [Giorgi, 1978] E. De Giorgi. Convergence problems for functionals or operators. *Proc. of the International Meeting on Recent Methods in Nonlinear Analysis, Pitagoria, Ed. Bologna*, 1978.
- [Giusti, 1984] E. Giusti. *Minimal surfaces and functions of bounded variation*. Birkhäuser, 1984.
- [Gobbino, 1998] M. Gobbino. Finite difference approximation of the Mumford-Shah functional. *Communication on pure and applied mathematics*, 51 :197–228, 1998.
- [Gomes et Faugeras, 1999] J. Gomes et O. Faugeras. Reconciling distance functions and level sets. In *Proc. of Scale-Space Theories in Computer Vision, Corfu, Grèce*, Septembre 1999.
- [Gomes et Faugeras, 2000] J. Gomes et O. Faugeras. Reconciling distance functions and level sets. *Journal of Visual Communication and Image Representation*, 11(2) :209–223, 2000.
- [Gonzales et Woods, 1992] R. Gonzales et R. Woods. *Digital image processing*. Addison-Wesley, 1992.
- [Graffigne *et al.*, 1995] C. Graffigne, F. Heitz, P. Pérez, F. Prêteux, M. Sigelle et J. Zerubia. Hierarchical markov random field models applied to image analysis : a review. In *Proc.*

- SPIE Conf. on neural, morphological, stochastic methods in image and signal, San Diego, Etats-Unis*, Juillet 1995.
- [Green, 1990] P.-J. Green. Bayesian reconstruction from emission tomography data using modified EM algorithm. *IEEE Trans. on Medical Imaging*, 9(1) :84–93, 1990.
- [Hadamard, 1923] J. Hadamard. *Lectures on Cauchy's problem in linear partial differential equations*. Yale University Press, 1923.
- [Hageman et Young, 1981] L. Hageman et D. Young. *Applied iterative methods*. Academic Press, 1981.
- [Hebert et Leahy, 1989] T. Hebert et R. Leahy. A generalized EM algorithm for 3D Bayesian reconstruction from Poisson data using Gibbs priors. *IEEE Trans. on Medical Imaging*, 8(2) :194–202, 1989.
- [Heitz et al., 1994] F. Heitz, P. Pérez et P. Bouthemy. Multiscale minimization of global energy functions in some visual recovery problems. *CVGIP : Image Understanding*, 59(1) :125–134, 1994.
- [Hidgon et al., 1997] D.M. Hidgon, J.E. Bowsher, V.E. Johnson, T.G. Turkington, D.R. Gilland et R.J. Jaszczak. Fully Bayesian estimation of Gibbs hyperparameters for emission computed tomography data. *IEEE Trans. on Medical Imaging*, 16(5) :516–526, Oct. 1997.
- [Jain et Farrokhnia, 1991] A. Jain et F. Farrokhnia. Unsupervised texture segmentation using Gabor filters. *Pattern recognition*, 24(12) :1167–1186, 1991.
- [Jain, 1989] A. Jain. *Fundamentals of digital image processing*. Prentice-Hall, 1989.
- [Jalobeanu et al., 1999] A. Jalobeanu, L. Blanc-Féraud et J. Zerubia. Hyperparameter estimation for satellite image restoration by a MCMCML method. In Springer, editor, *LNCS - EMMCVPR'99*, York, Grande Bretagne, Juillet 1999.
- [Kailath, 1967] T. Kailath. The divergence and Bhattacharyya distance measures in signal selection. *IEEE Trans. Communication Theory*, COM-15 :52–60, 1967.
- [Kass et al., 1987] M. Kass, A. Witkin et D. Terzopoulos. Snakes : active contour models. *International J. of Computer Vision*, 1 :321–331, 1987.
- [Kato et al., 1999] Z. Kato, J. Zerubia et M. Berthod. Unsupervised image classification using markovian models. *Pattern Recognition*, 32(4) :591–604, 1999.
- [Kato, 1994] Z. Kato. *Multiresolution Markovian modeling for computer vision. Application to SPOT image segmentation*. Thèse de Doctorat, Université de Nice-Sophia Antipolis, France, 1994.
- [Kichenassamy et al., 1996] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum et A. Yezzi Jr. Conformal curvature flows : from phase transitions to active vision. *Arch. Rational Mech. Anal.*, 134 :275–301, 1996.

-
- [Kittler, 1975] J. Kittler. Mathematical methods of feature selection in pattern recognition. *Int. Journal Man-Machine Studies*, 7 :609–637, 1975.
- [Kornprobst, 1998] P. Kornprobst. *Contribution à la restauration d'images et à l'analyse de séquences : approches variationnelles et solutions de viscosités*. Thèse de Doctorat, Université de Nice Sophia Antipolis, France, Novembre 1998.
- [Kreyszig, 1959] E. Kreyszig. *Differential geometry*. University of Toronto Press, Toronto, Canada, 1959.
- [Krishnapuram et Keller, 1994] R. Krishnapuram et J. M. Keller. *Fuzzy and possibilistic clustering methods for computer vision*, volume 12. SPIE institute series, in neural and fuzzy systems, 1994.
- [Krissian, 2000] K. Krissian. *Traitement multi-échelle : applications à l'imagerie médicale et à la détection tridimensionnelle de vaisseaux*. Thèse de Doctorat, Université de Nice Sophia Antipolis, France, Janvier 2000.
- [Laarhoven et Aarts, 1987] P. Van Laarhoven et E. Aarts. *Simulated Annealing : Theory and Applications*. Kluwer, 1987.
- [Laferté, 1996] J.-M. Laferté. *Contribution à l'analyse d'images par modèles markoviens sur des graphes hiérarchiques : application à la fusion de données multirésolutions*. Thèse de Doctorat, Université de Rennes I, France, 1996.
- [Lakshmanan et Derin, 1989] S. Lakshmanan et H. Derin. Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11 :799–813, Août 1989.
- [Lascaux et Theodor, 1987] P. Lascaux et R. Theodor. *Analyse numerique matricielle appliquee a l'art de l'ingenieur*. Masson, 1987.
- [Lazaroaia-Vese, 1996] L. Lazaroaia-Vese. *Problèmes variationnels et EDP pour l'analyse d'images et l'évolution de courbes*. Thèse de Doctorat, Université de Nice-Sophia Antipolis, France, 1996.
- [Li, 1995] S. Z. Li. *Markov Random Field Modeling in Computer Vision*. Springer-Verlag, 1995.
- [Lorette et al., 2000] A. Lorette, X. Descombes et J. Zerubia. Texture analysis through a Markovian modelling and fuzzy classification : application to urban area extraction from satellite images. *International Journal of Computer Vision*, 36(3) :221–236, 2000.
- [Lorette, 1999] A. Lorette. *Analyse de texture par méthodes Markoviennes et par morphologie mathématique : application à l'analyse des zones urbaines sur des images satellitaires*. Thèse de Doctorat, Université de Nice-Sophia Antipolis, France, 1999.

- [MacQueen, 1967] J. MacQueen. Some methods of classification and analysis of multivariate observations. In *Proc. 5th Berkeley Symposium on math. stat. and prob.*, Univ. of California Press, Berkeley, Etat-Unis, 1967.
- [Malladi *et al.*, 1994] R. Malladi, J.A. Sethian et B.C. Vemuri. Evolutionary fronts for topology independent shape modeling and recovery. In *Proc. of the 3rd ECCV*, pages 3–13, Stockholm, Suède, 1994.
- [Malladi *et al.*, 1995] R. Malladi, J.A. Sethian et B.C. Vemuri. Shape modeling with front propagation : a level set approach. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(2) :158–175, 1995.
- [Manjunath et Chellappa, 1991] B. Manjunath et R. Chellappa. Unsupervised texture segmentation using Markov random fields models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 13 :478–482, Mai 1991.
- [March et Dozio, 1997] R. March et M. Dozio. A variational method for the recovery of smooth boundaries. *Image and Vision Computing*, 15 :705–712, 1997.
- [March, 1992] R. March. Visual reconstruction using variational methods. *Image and Vision Computing*, 10 :30–38, 1992.
- [Marroquin *et al.*, 1987] J. Marroquin, S. Mitter et T. Poggio. Probabilistic solution of ill-posed problems in computational vision. *Journal of the american statistical association*, 82 :76–89, 1987.
- [Maso, 1992] G. Dal Maso. *Introduction to Γ -convergence*. Birkhäuser, 1992.
- [Merriman *et al.*, 1994] B. Merriman, J. Bence et S. Osher. Motion of multiple junctions : a level set approach. *J. of computational physics*, 112 :334–363, 1994.
- [Metropolis *et al.*, 1953] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller et E. Teller. Equations of state calculations by fast computing machines. *Journal of chemical physics*, 21 :1087–1091, 1953.
- [Modica, 1987] L. Modica. The gradient theory of phase transitions and the minimal interface criterion. *Arch. Rational Mech. Anal.*, 98 :123–142, 1987.
- [Morel et Solimini, 1995] J.-M. Morel et S. Solimini. *Variational methods in image segmentation*. Birkhäuser, 1995.
- [Morris *et al.*, 1996] R. Morris, X. Descombes et J. Zerubia. The Ising/Potts model is not well suited to segmentation tasks. In *proc. IEEE Digital Signal Processing Workshop*, 1996. Sept. 1-4 1996 , Norvège.
- [Mumford et Shah, 1985] D. Mumford et J. Shah. Boundary detection by minimizing functionals. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, San Francisco, Etats-Unis, 1985.

-
- [Nelder et Mead, 1965] J. Nelder et R. Mead. A simplex method for function minimization. *The Computer Journal*, 7, 1965.
- [Nitzberg et Shiotani, 1992] M. Nitzberg et T. Shiotani. Nonlinear image filtering with edge and corner enhancement. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(8) :826–833, 1992.
- [Nordström, 1990] K.N. Nordström. Biased anisotropic diffusion : a unified regularization and diffusion approach to edge detection. *Image and Vision Computing*, 8(4) :318–327, 1990.
- [Osher et Sethian, 1988] S. Osher et J.A. Sethian. Fronts propagating with curvature dependent speed : algorithms based on the Hamilton-Jacobi formulation. *J. of Computational Physics*, 79 :12–49, 1988.
- [Owen et Sternberg, 1991] C. Owen et P. Sternberg. Nonconvex variational problems with anisotropic perturbations. *Nonlinear Analysis, Theory, Methods and Applications*, 16(7) :705–719, 1991.
- [Paragios et Deriche, 1998] N. Paragios et R. Deriche. A PDE-based level set approach for detection and tracking of moving objects. In *Proc. ICCV, Bombay, Inde*, pages 1139–1135, 1998.
- [Paragios et Deriche, 1999] N. Paragios et R. Deriche. Geodesic active regions for supervised texture segmentation. In *Proc. of ICCV, Corfu, Grèce*, pages 926–932, Septembre 1999.
- [Paragios, 2000] N. Paragios. *Geodesic active regions and level set methods : contributions and applications in artificial vision*. Thèse de Doctorat, Université de Nice Sophia Antipolis, France, Janvier 2000.
- [Pauc et al., 1996] G. Pauc, C. Latry, A. Pasquet et E. Breton. Pourquoi et comment simuler des images SPOT 5. *Bul. S.F.P.T.*, 141 :22–27, janvier 1996.
- [Pavlidis et Liow, 1988] T. Pavlidis et Y.-T. Liow. Integrating region growing and edge detection. In *Proc. of IEEE CVPR*, 1988.
- [Perona et Malik, 1990] P. Perona et J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(7) :629–639, 1990.
- [Pérez et Heitz, 1996] P. Pérez et F. Heitz. Restriction of a markov random field on a graph and multiresolution statistical image modeling. *IEEE Trans. Information Theory*, 42(1) :180–190, 1996.
- [Pérez, 1993] P. Pérez. *Champs markoviens et analyse multirésolution : application à l'analyse du mouvement*. Thèse de Doctorat, Université de Rennes I, France, 1993.

- [Pérez, 1998] P. Pérez. Markov random fields and images. *CWI Quarterly*, 11(4) :413–437, 1998.
- [Reitich et Soner, 1996] F. Reitich et H.M. Soner. Three-phase boundary motions under constant velocities. part one : The vanishing surface tension limit. *Proc. of the Royal Society of Edinburgh*, 126(A) :837–865, 1996.
- [Richards, 1993] J. Richards. *Remote sensing digital image analysis*. Springer-Verlag, deuxièm ed., 1993.
- [Romeny, 1994] Bart M. Ter Haar Romeny. *Geometry-driven diffusion in computer vision*. Kluwer academic, 1994.
- [Rostaing *et al.*, 1999] P. Rostaing, J.-N. Provost et C. Collet. Unsupervised multispectral image segmentation using generalized Gaussian noise model. In *Int. Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition, York, G.B.*, Juillet 1999.
- [Rubinstein *et al.*, 1989] J. Rubinstein, P. Sternberg et J.B. Keller. Fast reaction, slow diffusion, and curve shortening. *SIAM J. of Applied Mathematics*, 49 :116–133, 1989.
- [Rudin *et al.*, 1992] L. Rudin, S. Osher et E. Fatemi. Nonlinear total variation based removal algorithm. *Physica D*, 60 :259–268, 1992.
- [Rykov, 1980] A. Rykov. Simplex direct search algorithms. *Automation and Remote Control*, 41, 1980.
- [Samson *et al.*, 1998] C. Samson, L. Blanc-Féraud, G. Aubert et J. Zerubia. Image classification using a variational approach. Rapport de Recherche INRIA RR-3523, (<http://www.inria.fr/RRRT/RR-3523.html>), Octobre 1998.
- [Samson *et al.*, 1999a] C. Samson, L. Blanc-Féraud, G. Aubert et J. Zerubia. A level set model for image classification. In *Proc. of Scale-Space Theories in Computer Vision, Corfu, Grèce*, Septembre 1999.
- [Samson *et al.*, 1999b] C. Samson, L. Blanc-Féraud, G. Aubert et J. Zerubia. Classification et restauration d’images par approche variationnelle. In *Proc. ORASIS*, Avril 1999.
- [Samson *et al.*, 1999c] C. Samson, L. Blanc-Féraud, G. Aubert et J. Zerubia. Multiphase evolution and variational image classification. Rapport de Recherche INRIA RR-3662, (<http://www.inria.fr/RRRT/RR-3662.html>), Avril 1999.
- [Samson *et al.*, 1999d] C. Samson, L. Blanc-Féraud, G. Aubert et J. Zerubia. Simultaneous image classification and restoration using a variational approach. In *Proc. of IEEE CVPR, Fort-Collins, Etats-Unis*, Juin 1999.

-
- [Samson *et al.*, 2000a] C. Samson, L. Blanc-Féraud, G. Aubert et J. Zerubia. A level set model for image classification. *à paraître dans International Journal of Computer Vision*, 2000.
- [Samson *et al.*, 2000b] C. Samson, L. Blanc-Féraud, G. Aubert et J. Zerubia. A variational model for image classification and restoration. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(5) :460–472, Mai 2000.
- [Sapiro et Ringach, 1996] G. Sapiro et D.L. Ringach. Anisotropic diffusion of multivalued images with applications to color filtering. *IEEE Trans. on Image Processing*, 5(11) :1582–1586, Novembre 1996.
- [Sethian, 1996] J. Sethian. *Level-Set method*. Cambridge University Press, 1996.
- [Sethian, 1999] J. Sethian. *Level Set Methods and Fast Marching Methods*. Cambridge University Press, 1999.
- [Sternberg et Zeimer, 1994] P. Sternberg et W.P. Zeimer. Local minimisers of a three-phase partition problem with triple junctions. *Proc. of the Royal Society of Edinburgh*, 124(A) :1059–1073, 1994.
- [Sternberg, 1991] P. Sternberg. Vector-valued local minimizers of nonconvex variational problems. *J. of Mathematics*, 21 :799–807, 1991.
- [Sussman *et al.*, 1994] M. Sussman, P. Smereka et S. Osher. A level set approach for computing solutions to incompressible two-phase flow. *J. of Computational Physics*, 114 :146–159, 1994.
- [Teboul *et al.*, 1998] S. Teboul, L. Blanc-Féraud, G. Aubert et M. Barlaud. Variational approach for edge-preserving regularization using coupled PDE's. *IEEE Trans. on Image Processing*, 7(3) :387–397, 1998.
- [Teboul, 1999] S. Teboul. *Reconstruction et Segmentation d'Images : Approche Variationnelle et EDP couplées*. Thèse de Doctorat, Université de Nice-Sophia Antipolis, France, Avril 1999.
- [Tikhonov et Arsenin, 1977] A.N. Tikhonov et V.Y. Arsenin. *Solutions of ill-posed problems*. Winston and Wiley, 1977.
- [Torre et Poggio, 1986] V. Torre et T. Poggio. On edge detection. *IEEE Trans. on Pattern analysis and machine intelligence*, 8(3) :147–163, 1986.
- [Volden *et al.*, 1996a] E. Volden, G. Giraudon et M. Berthod. Information in Markov random fields and image redundancy. In *Lecture Notes in Computer Science : information theory and applications*, Septembre 1996.
- [Volden *et al.*, 1996b] E. Volden, G. Giraudon et M. Berthod. La redondance d'images : étude et application. In *Proc. RFIA, Rennes, France*, Janvier 1996.

- [Weickert, 1998] J. Weickert. *Anisotropic Diffusion in Image Processing*. ECMI Series, Teubner-Verlag, 1998.
- [Whitaker et Gerig, 1994] R. Whitaker et G. Gerig. *Vector-valued diffusion*. computational imaging and vision : geometry driven diffusion in computer vision, Kluwer, 1994.
- [Winter, 1997] A. Winter. *Entropie et représentations multi-échelles pour l'interprétation automatique d'images satellitaires et aériennes*. Thèse de Doctorat, Ecole Nationale Supérieure des Télécommunications, France, 1997.
- [Wolberg et Pavlidis, 1985] G. Wolberg et T. Pavlidis. Restoration of binary images using stochastic relaxation with annealing. *Pattern Recognition Letters*, 3 :375–388, 1985.
- [Wu, 1982] F. Wu. The Potts model. *Reviews of modern physics*, 54(1) :235–268, 1982.
- [Yezzi *et al.*, 1999] A. Yezzi, Jr., A. Tsai et A. Willsky. A statistical approach to snakes for bimodal and trimodal imagery. In *Proc. of ICCV, Corfu, Grèce*, pages 898–903, Septembre 1999.
- [You *et al.*, 1996] Y. You, W. Xu, A. Tannenbaum et M. Kaveh. Behavioral analysis of anisotropic diffusion in image processing. *IEEE Trans. on Image Processing*, 5(11) :1539–1553, 1996.
- [You et Kaveh, 1999] Y. You et M. Kaveh. Blind image restoration by anisotropic regularization. *IEEE Trans. on Image Processing*, 8(3) :396–407, Mars 1999.
- [Zadeh, 1965] L. A. Zadeh. Fuzzy sets. *Information and control*, 8 :338–353, 1965.
- [Zenko, 1986] S. Di Zenzo. A note on the gradient of a multi-image. *CVGIP*, 33 :116–125, 1986.
- [Zhao *et al.*, 1996] H-K. Zhao, T. Chan, B. Merriman et S. Osher. A variational level set approach to multiphase motion. *J. of Computational Physics*, 127 :179–195, 1996.
- [Zhu et Yuille, 1996] S. C. Zhu et A. Yuille. Integrating region growing and edge detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(9) :884–900, 1996.

Publications de l'auteur

Les travaux effectués au cours de cette thèse ont donné lieu aux publications suivantes :

Journaux

- C. Samson, L. Blanc-Féraud, G. Aubert et J. Zerubia. Deux Modèles Variationnels pour la Classification d'Images Multispectrales. Soumis à *Traitement du Signal*, Septembre 2000.
- C. Samson, L. Blanc-Féraud, G. Aubert et J. Zerubia. A Level Set Model for Image Classification. A paraître dans *International Journal of Computer Vision*, 2000.
- C. Samson, L. Blanc-Féraud, G. Aubert et J. Zerubia. A variational model for image classification and restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(5) : 460-472, Mai 2000.

Conférences internationales

- C. Samson, L. Blanc-Féraud, G. Aubert et J. Zerubia. A level set model for image classification. *Proc. of Scale-Space Theories in Computer Vision*, Corfu, Grèce, Septembre 1999.
- C. Samson, L. Blanc-Féraud, G. Aubert et J. Zerubia. Simultaneous image classification and restoration using a variational approach. *Proc. IEEE Computer Vision and Pattern Recognition*, Fort-Collins, Etats-Unis, Juin 1999.
- C. Samson, L. Blanc-Féraud, G. Aubert et J. Zerubia. Image classification using a variational approach. *Mathématiques et Reconnaissance de Formes*, CIRM Luminy, France, Avril 1999.

Conférences nationales

- C. Samson, L. Blanc-Féraud, G. Aubert et J. Zerubia. Une approche variationnelle pour la classification d'images par régions actives. *Reconnaissance de Formes et Intelligence Artificielle*, Paris, Février 2000.

- C. Samson, L. Blanc-Féraud, G. Aubert et J. Zerubia. Classification et Restauration d'Images par Approche Variationnelle. *ORASIS*, Aussois, Avril 1999.

Rapports de recherche

- C. Samson, L. Blanc-Féraud, G. Aubert et J. Zerubia. Classification d'Images Multibandes par Modèles Variationnels. *Rapport de Recherche INRIA RR-4010*, Septembre 2000.
- C. Samson, L. Blanc-Féraud, G. Aubert et J. Zerubia. Multiphase evolution and variational image classification. *Rapport de Recherche INRIA RR-3662*, Avril 1999.
- C. Samson, L. Blanc-Féraud, G. Aubert et J. Zerubia. Image classification using a variational approach. *Rapport de Recherche INRIA RR-3523*, Octobre 1998.

RÉSUMÉ

Ce travail est consacré au développement ainsi qu'à l'implantation de deux modèles variationnels pour la classification d'images.

La classification d'images, consistant à attribuer une étiquette à chaque pixel d'une image, concerne de nombreuses applications à partir du moment où cette opération intervient très souvent à la base des chaînes de traitement et d'interprétation d'images. De nombreux modèles de classification ont déjà été développés dans un cadre stochastique ou à travers des approches structurales, mais rarement dans un contexte variationnel qui a déjà montré son efficacité dans divers domaines tels que la reconstruction ou la restauration d'images.

Le premier modèle que nous proposons repose sur la minimisation d'une famille de critères dont la suite de solutions converge vers une partition des données composée de classes homogènes séparées par des contours réguliers. Cette approche entre dans le cadre des problèmes à discontinuité libre (*free discontinuity problems*) et fait appel à des notions de convergence variationnelle telle que la théorie de la Γ -convergence. La famille de fonctionnelles que nous proposons de minimiser contient un terme de régularisation, ainsi qu'un terme de classification. Lors de la convergence de cette suite de critères, le modèle change progressivement de comportement en commençant par restaurer l'image avant d'entamer le processus d'étiquetage des pixels.

Parallèlement à cette approche, nous avons développé un second modèle de classification mettant en jeu un ensemble de régions et contours actifs. Nous utilisons une approche par ensembles de niveaux pour définir le critère à minimiser, cette approche ayant déjà suscité de nombreux travaux dans le cadre de la segmentation d'images. Chaque classe, et son ensemble de régions et contours associé, est défini à travers une fonction d'ensemble de niveaux. Le critère contient des termes reliés à l'information sur les régions ainsi qu'à l'information sur les contours. Nous aboutissons à la résolution d'un système d'équations aux dérivées partielles couplées et plongées dans un schéma dynamique. L'évolution de chaque région est guidée par un jeu de forces permettant d'obtenir une partition de l'image composée de classes homogènes et dont les frontières sont lisses.

Nous avons mené des expériences sur de nombreuses données synthétiques ainsi que sur des images satellitaires SPOT. Nous avons également étendu ces deux modèles au cas de données multispectrales et obtenu des résultats sur des données SPOT XS que nous avons comparé à ceux obtenus par différents modèles.

Mots clefs : classification, restauration, Γ -convergence, ensembles de niveaux, régions et contours actifs, multispectral, imagerie aérienne et satellitaire.

ABSTRACT

This work is devoted to the development and the implementation of variational models for image classification.

Image classification, which consists in assigning a label to each pixel of a given image, concerns many applications since it is often the basic processing for many image interpretation systems. Many models have been developed within a stochastic framework or using structural approaches, but rarely within a variational framework whose efficiency has largely been proved for a wide variety of problems such as image reconstruction or restoration.

The first model we propose herein is based on the minimization of a criterion family whose set of solutions in converging to a partition of the data set composed of homogeneous regions with regularized boundaries. This approach takes place within the context of free boundary problems and we use the Γ -convergence theory for the theoretical study. The set of functionals we minimize contains a regularization term and a classification one. As the set of functionals is converging, the behavior of the model is progressively changing : the restoration process is vanishing while the labeling one is rising.

The second model we propose is based on a set of active regions and contours. We use a level set formulation to define the criterion we want to minimize, this formulation allows a change of topology of the evolving sets. Each class and its associated set of regions and boundaries is defined thanks to a level set function. From the Euler equations, we solve a system of coupled partial differential equations through a dynamical scheme. The evolution of each region is governed by forces constraining the partition to be composed of homogeneous classes with smooth boundaries.

We have conducted many experiments on both synthetic and real images. We have extended these models to the multispectral case for which the data are a set of images, and we show some results and comparisons on SPOT XS images.

Keywords : Classification, restoration, Γ -convergence, level set, active regions and contours, multispectral, aerial and satellite imaging